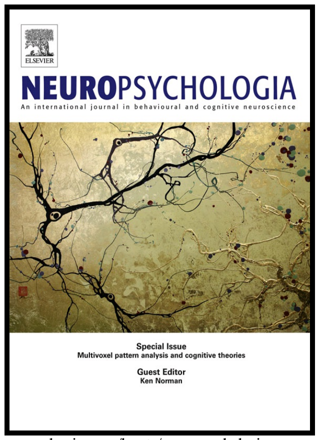
Author's Accepted Manuscript



www.elsevier.com/locate/neuropsychologia

Neural Mechanisms underlying Valence Inferences to Sound: the Role of the right Angular Gyrus

Fernando Bravo, Ian Cross, Sarah Hawkins, Nadia Gonzalez, Jorge Docampo, Claudio Bruno, Emmanuel Andreas Stamatakis

PII: S0028-3932(17)30204-X
DOI: http://dx.doi.org/10.1016/j.neuropsychologia.2017.05.029
Reference: NSY6380

To appear in: *Neuropsychologia*

Received date: 22 July 2016
Revised date: 24 May 2017
Accepted date: 31 May 2017

Cite this article as: Fernando Bravo, Ian Cross, Sarah Hawkins, Nadia Gonzalez Jorge Docampo, Claudio Bruno and Emmanuel Andreas Stamatakis, Neura Mechanisms underlying Valence Inferences to Sound: the Role of the righ Angular Gyrus, *Neuropsychologia* http://dx.doi.org/10.1016/j.neuropsychologia.2017.05.029

Neural Mechanisms underlying Valence Inferences to Sound: the Role of the right Angular Gyrus.

Fernando Bravo[1,2*], Ian Cross[1], Sarah Hawkins[1], Nadia Gonzalez[3], Jorge Docampo[3], Claudio Bruno[3], Emmanuel Andreas Stamatakis[4]

[1]University of Cambridge. Centre for Music and Science (UK, Cambridge).

[2]TU Dresden. Institut für Kunst- und Musikwissenschaft (Germany, Dresden).

[3]Fundación Científica del Sur Imaging Centre (Argentina, Buenos Aires).

[4]University of Cambridge. Division of Anaesthesia (UK, Cambridge).

Institutions at which the work was performed

University of Cambridge and TU Dresden.

*Corresponding author. Dr. Fernando Bravo, Institut für Kunst- und Musikwissenschaft Technische Universität Dresden August-Bebel-Straße 20, 01219 Dresden, Germany. Telephone: +4917626699400. nanobravo@fulbrightmail.org

Abstract

We frequently infer others' intentions based on non-verbal auditory cues. Although the brain underpinnings of social cognition have been extensively studied, no empirical work has yet examined the impact of musical structure manipulation on the neural processing of emotional valence during mental state inferences. We used a novel sound-based theory-of-mind paradigm in which participants categorized stimuli of different sensory dissonance level in terms of positive/negative valence. Whilst consistent with previous studies which propose facilitated encoding of consonances, our results demonstrated that distinct levels of consonance/dissonance elicited differential influences on the right angular gyrus, an area implicated in mental state attribution and attention reorienting processes. Functional and effective connectivity analyses further showed that consonances modulated a specific inhibitory interaction from associative memory to mental state attribution substrates. Following evidence suggesting that individuals with autism may process social affective cues differently, we assessed the relationship between participants' task performance and self-reported autistic traits in clinically typical adults. Higher scores on the social cognition scales of the AQ were associated with deficits in recognising positive valence in consonant sound cues. These findings are discussed with respect to Bayesian perspectives on autistic perception, which highlight a functional failure to optimize precision in relation to prior beliefs.

Keywords: Consonance/Dissonance, Angular Gyrus, Music, Emotion, Theory of Mind.

Introduction

To navigate the social environment, humans rely on their capacity to recognise cues signalling potentially threatening or affiliative value in others' mental states. The present study builds upon knowledge derived from music perception to elucidate cognitive mechanisms and neural systems involved in this ability.

The proposition that pleasant-sounding (consonant) combinations of tones entail special numerical properties has been ascribed to Pythagoras (Apel, 1972). He is supposed to have observed that tones produced by partitioning a vibrating string in two segments with lengths related by simple (i.e. small-integer) ratios, such as 2:1, 3:2 and 4:3, resulted in more pleasing harmonies compared to those produced by more complex ratios (e.g. 9:8, 16:15). Empirical evidence from studies conducted with infants, children, and adults, indeed suggests that sequential pure-tone intervals with simple frequency ratios confer perceptual processing advantages (Schellenberg and Trehub, 1994, 1996a, 1996b). Researchers have argued that intervals with simple ratios would be inherently easier to encode, manage and recognise as a unit (i.e. more coherent: Frances, 1972; Bharucha and Prior, 1996) forming prototypes (Rosch, 1975) that would provide a perceptual frame of reference for distinguishing other intervals (Trehub and Unyk, 1991). It has been proposed that the special perceptual status of intervals with simple frequency ratios, such as the octave (2:1), perfect fifth (3:2), and perfect fourth (4:3), could stem from their presence in naturally occurring sounds including those of speech and music (Terhardt, 1974a, 1978, 1984), or that it may result from the exposure to particular musical cultures or styles (Serafine, 1983; Downling and Harwood, 1986). Such distinctiveness has also been reflected in judgments of consonance and dissonance; with simple-ratio intervals being judged more consonant (i.e. more pleasant, smooth and well blended) than intervals with more complex ratios, such as the major second (9:8), minor second (16:15), and tritone (45:32), which have been consistently evaluated as more dissonant (i.e. more unpleasant and less smooth) (Schellenberg and Trehub, 1994; Plomp and Levelt, 1965; Wedin, 1972; Zentner and Kagan, 1996; Schellenberg and Trainor, 1996; Trainor and Heinmiller, 1998; Blood et al. 1999).

The present study focuses on the emotional effects and, in particular, on the valence judgments elicited by musical intervals of different degrees of consonance/dissonance. There is substantial evidence showing that level of consonance/dissonance is strongly associated with the percept of valence (Blood et al., 1999; Costa et al., 2000; Plomp and Levelt, 1965; Trainor and Heinmiller, 1998). Valence has been defined as the subjective feeling of pleasantness or unpleasantness (Barrett and Wager, 2006; Lindquist et al., 2012; Russell, 1979). With regards to social interaction, valence has been conceptualized as the intrinsic attractiveness/good-ness (positive valence) or averseness/bad-ness (negative valence) of an event, object or situation (Colombetti, 2005; Frijda, 1986). Together with arousal and potency, they have been proposed as the three affective dimensions widely considered to explain the fundamental variance of emotional responses (Lang et al., 1993; Russell, 1979). Researchers have frequently utilized the valence percept as an indirect measure to assess degree of consonance/dissonance (Blood et al., 1999; Gosselin et al., 2006; Koelsch et al., 2006; Plomp and Levelt, 1965; Trainor and Heinmiller, 1998). Valence inferences have been shown to consistently index the perception of consonance/dissonance level in Western musicians and non-musicians (Blood et al., 1999; Bugg, 1970; Plomp and Levelt, 1965) and an association between valence and degree of dissonance has been also reported in listeners never exposed to Western music (Fritz et al., 2009). Although the affective appraisal of musical dissonance seems to be strongly influenced by culture, as demonstrated by studies that have documented its variations across different cultures and its historical transformation through distinct Western culture periods (Burns, 1999), valence judgments applied to stimuli with distinct degrees of sensory dissonance (the type of dissonance manipulated in the present study) appear to be culturally invariant and largely independent of musical training [(Bidelman and Krishnan, 2011; Chiandetti and Vallortigara, 2011; Fannin and Braud, 1971; Foss et al., 2007; Fujisawa and Cook, 2011; Itoh et al., 2010; Izumi, 2000;

Minati et al., 2009; Peretz et al., 2001; Sugimoto et al., 2010; Zentner and Kagan, 1996) although see also: (McDermott and Hauser, 2004; McDermott et al., 2016)].

Various psychoacoustic models have been suggested to elucidate why musical intervals comprising simple frequency ratios are experienced as more consonant than intervals involving complex ratios (Helmholtz and Ellis, 1895; Kameoka and Kuriyagawa, 1969; Plomp and Levelt, 1965; Terhardt, 1978). One influential theory was coined by Helmholtz (1895), who proposed that sensory consonance/dissonance was related to the absence/presence of interactions (sensation of “beats” or “roughness”) between the harmonic spectra of two pitches. Empirical evidence has also shown that the perception of consonance/dissonance can be elicited not only by the properties of a single signal, such as roughness, but also when tones are presented dichotically (i.e. when different pitches are presented separately to each ear) (Cousineau et al., 2012; Fritz et al., 2013; McDermott et al., 2010). Fritz and collaborators (2013) have shown that dichotic dissonance stimulation also elicits negative valence ratings, which indicates that cochlear interactions may not be critical for the perception of dissonance. It is important to note, however, that during dichotic listening tasks, the allocation of attention in the auditory space can be modulated by training (Soveri et al., 2013) and, consequently, participants’ valence judgments during dichotic paradigms could also be explained by attentional focus on one ear. To overcome this potential problem, in the present work we employed sequential intervals presented diotically (each tone was audible by both ears simultaneously), which do not produce roughness or beats due to their non-simultaneity; yet sequential intervals are also known to be judged along the dimension of consonance/dissonance according to their frequency ratios (Ayres et al., 1980; Fritz et al., 2013; Schellenberg and Trehub, 1994).

Several studies have investigated the neural correlates of emotional responses to dissonance. Five relevant neuroscientific studies should be mentioned. The study by Blood and collaborators (1999) used positron emission tomography (PET) to measure the brain correlates of negative affective reactions induced by dissonance. Degree of dissonance level was controlled by presenting participants with a novel melody, which was manipulated through altering the harmonic structure of its accompanying chords. A preliminary behavioural study showed that higher levels of dissonance were correlated with higher average ratings of adjectives associated with negative emotions (e.g. tense, unpleasant, irritated, annoying, dissonant and angry). Participants were informed that the experimenters were interested in their emotional responses to music, and they were asked to respond to an emotional discrimination task (rating emotional valence). Increasing dissonance correlated with activity in right parahippocampal gyrus and right precuneus. Higher ratings of unpleasantness, which correlated with increasing dissonance, covaried with cerebral blood flow changes in right parahippocampal gyrus and left posterior cingulate. On the other hand, activations in orbitofrontal, subcallosal cingulate and frontal polar cortex correlated with decreasing dissonance (equivalent to increasing consonance). Koelsch et al. (2006) used fMRI to investigate the brain circuits mediating emotions with positive and negative valence elicited by consonant and permanently dissonant counterparts of the original tunes (classical music from the common practice period). In contrast to the study by Blood et al. (1999), which used musical stimuli presented via computerized control and without musical expression, they employed naturalistic music taken from commercially available CDs. The unpleasant stimuli were obtained by electronically manipulating "joyful" naturalistic instrumental dance-tunes ('the original -pleasant- excerpt was recorded simultaneously with two pitch shifted versions of the same excerpt, the pitch-shifted versions being one tone above and a tritone below the original pitch'). Participants had to indicate how pleasant or unpleasant they felt following each musical excerpt. During the presentation of unpleasant music (contrasted to pleasant music), activations were found in the left hippocampus, the left parahippocampal gyrus, the right temporal pole, and the left amygdala. When contrasting pleasant vs. unpleasant music, they observed activations of Heschl's gyrus, the anterior superior insula, and the left inferior frontal gyrus. In the study by Gosselin et al. (2006), a group of epileptic patients with anteromedial lobe excision were examined (brain regions removed included variable amounts of parahippocampal, perihinal, entorhinal and hippocampal tissue). Patients were asked to rate the degree of pleasantness of consonant and dissonant manipulated versions of the same happy or sad musical excerpts. They noticed that patients with parahippocampal resection showed diminished sensitivity to unpleasant (dissonant) music, judging the dissonant stimuli as moderately pleasant (significantly happier compared to normal controls). The authors interpreted the contribution of the parahippocampal cortex as 'specific to the emotional interpretation of dissonance'. A fourth relevant study, conducted by Green et al. (2008), which was aimed at exploring brain activity underlying musical mode perception, found increased activity in the left parahippocampal gyrus, bilateral ventral anterior cingulate cortex and left medial prefrontal cortex in response to minor mode melodies, compared to equivalent major melodies. The authors proposed harmonic dissonance as a possible contributing factor in the observed minor related activity increase. Finally, also using fMRI Foss et al. (2007) found that the anterior cingulate cortex, inferior frontal gyrus, superior temporal gyrus, medial frontal gyrus and inferior parietal lobule responded to increasing dissonance with progressively more activation.

Although the findings of these studies only partially overlap (and in some cases contradict each other), they converge in supporting a critical role of the parahippocampal cortex and, to a less extent, of medial prefrontal cortices (e.g. anterior cingulate cortex and medial prefrontal cortex) in the emotional evaluation of perceived degrees of dissonance (Blood et al., 1999; Foss et al., 2007; Gosselin et al., 2006; Green et al., 2008; Koelsch et al., 2006).

The involvement of parahippocampal and medial prefrontal cortices has been also supported by studies that have examined the processing of other emotional aspects of music (Mitterschiffthaler et al., 2007). In particular, the evaluation of emotional content (including valence judgments) coded in both voice and music (Escoffier et al., 2013) as well as during the processing of voice, body movements and facial expressions (Peelen et al., 2010) has been found to recruit the medial prefrontal cortex, reflecting an important role of these regions in encoding supramodal emotion representations. Moreover, the fact that the medial prefrontal cortex is a core region of the mentalizing or theory of mind (ToM) network has prompted researchers to propose a common reliance on processes involved in social cognition (Fletcher et al., 1995; Gallagher et al., 2000; Gobbini et al., 2007; Saxe and Wexler, 2005).

The degree of consonance/dissonance has been found to influence liking and the perception of emotional content not only in music (Blood et al. 1999; Koelsch et al. 2006) but also in audiovisual contexts (Cohen, 2001; Boltz, 2001; Bravo, 2012, 2014). Connotations ascribed to consonant and dissonant sounds have been frequently employed in film sound to shape the emotional comprehension of visual narratives, mainly through influencing the attribution of mental states (emotions, thoughts or intentions) to the characters depicted onscreen (Cohen, 2001; Boltz, 2001; Bravo, 2012, 2014). Notably, to our best knowledge, no empirical work has yet examined the impact of consonance/dissonance on the neural substrates underlying valence inference processes during mental state attribution.

In the present study, information conveyed through sound acted as the non-verbal cue to be interpreted. Subjects had to make temporary valence inferences (Van Overwalle et al., 2009) based on auditory signals that only differed in terms of consonance/dissonance level, which was controlled by interval content manipulation. The task employed a purposely-made metaphor, which informed the participants that a radio-telescope had captured a series of radio signals from outer space. Participants were asked to listen to these signals, and to think and decide if they were produced by good-friendly or bad-aggressive aliens (Figure 1, left). The task therefore required to *predict the affective value of a message conveyed via musical intervals*. In the present study, participants' judgements were considered as inferences of transitory states or temporary inferences (Van Overwalle, 2009; Van Overwalle et al., 2009), which have been found to rely on theory of mind function (Abell et al., 2000; Castelli et al., 2002; Martin and Weisberg, 2003; Saxe and Wexler, 2005; Schultz et al., 2004; Van Overwalle et al., 2009).

The first two experiments reported in the present article -behavioural, and fMRI- investigated the effects of consonance/dissonance level on the cognitive and neural mechanisms underlying the valence judgments during temporary inferences ascribed to an agent (e.g. is he or she friendly or aggressive?) (Van Overwalle, 2009).

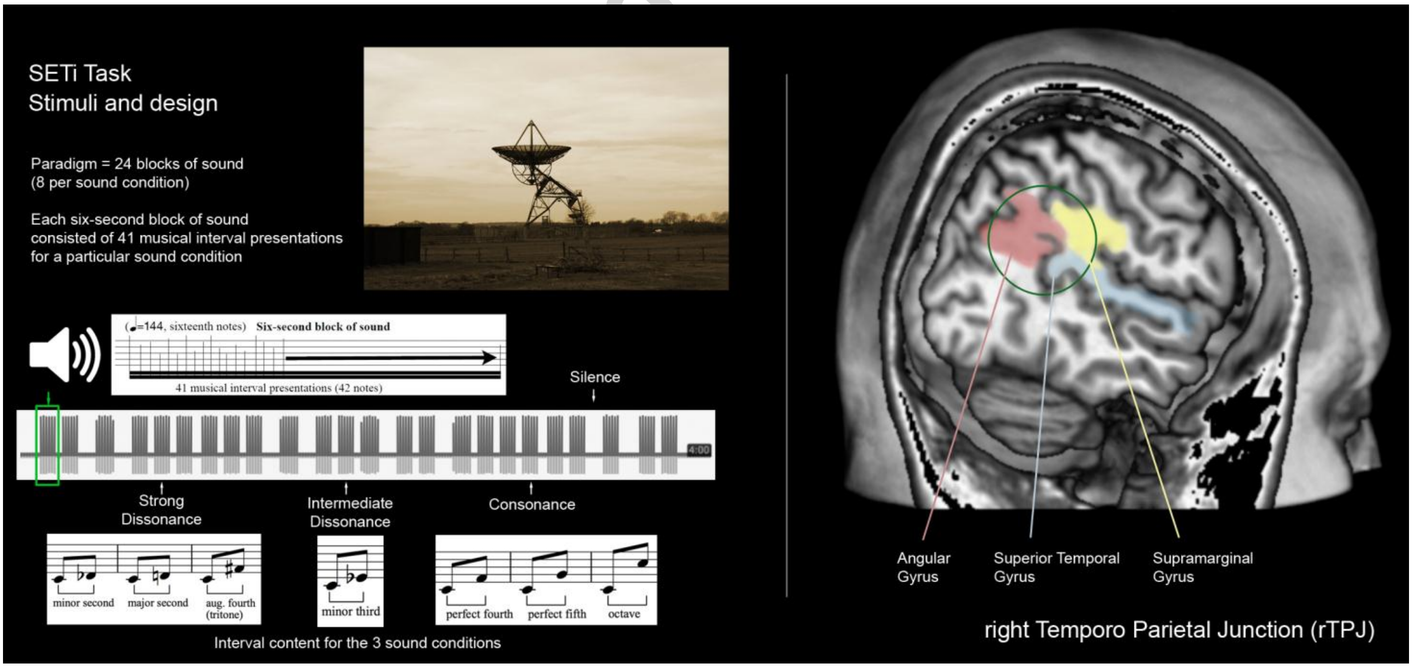


Fig. 1. (left) SETi (Search for Extra-Terrestrial integrity) task. Subjects viewed the above image of a radio-telescope and were given the following instruction: "A radio-telescope located in Cambridge captured a series of radio signals from outer space. You will listen to these sounds and your task is to think and decide if they were produced by good-friendly or bad-aggressive aliens".

(right) The right temporo-parietal junction (rTPJ) is located within the green outline. It comprises portions of the angular gyrus (AG, red), supramarginal gyrus (SMG, yellow), and superior temporal gyrus (STG, blue).

The paradigm required the ability to attribute mental states in real-time, also referred to as “on-line” mentalizing (Abell et al., 2000; Castelli et al., 2002). To date, researchers have extensively examined the brain structures that are engaged when performing tasks that require the attribution of mental states. A network has been identified, comprising midline cortical structures (medial prefrontal, anterior and posterior cingulate cortices) as well as the bilateral temporo-parietal junction (Frith and Frith, 2003; Schurz and Perner, 2015; Marchetti et al. 2015). In particular, there is strong evidence suggesting that the right temporo-parietal junction implements intuitive and empathic representations about *temporary intentions attributed to an actor* (Keysers and Perrett, 2004; Saxe and Wexler, 2005; Keysers and Gazzola, 2007; Van Overwalle, 2009; Van Overwalle and Baetens, 2009), whilst the medial prefrontal cortex (mPFC) is implicated in identifying enduring traits, which involve more reflective representations and deliberate reasoning (Mitchell et al., 2005, 2006; Todorov et al., 2007). The temporo-parietal junction is an area of the cerebral cortex situated along the boundary between the temporal and parietal lobes that plays a critical role in various aspects of social cognition, which have in turn been proposed to rely on lower-level self-other discrimination processes (Decety and Lamm, 2007; Santiesteban et al., 2012; Silani et al., 2013). It encompasses portions of the angular gyrus, supramarginal gyrus, and superior temporal gyrus and sulcus (Figure 1, right). Importantly, activation in the right angular gyrus seems to be strongly linked to processes requiring intentional attribution and temporary inferences (rTPJ-Mental: Saxe, 2006, 2010, Scholz et al. 2009; Carter and Huettel, 2013; Van Overwalle, 2009). We therefore hypothesized that our task would recruit right temporo-parietal junction areas (rTPJ) and, specifically entail processing in the right angular gyrus (rAG).

We predicted that distinct consonance/dissonance levels would exert differential influences on mental state attribution substrates during valence inferences. Based on predictive coding models (see Box 1), which posit that higher areas (e.g. memory systems) are actively attempting to “explain” incoming information represented in lower areas via feedback projections (Rao and Ballard 1999, Friston, 2009), we hypothesised that the perceptual processing advantages consistently demonstrated for consonant intervals (i.e. more coherent structure: Schellenberg and Trehub, 1994, 1996a, 1996b) would be reflected by increased activity in mental state attribution substrates during valence judgements for dissonances compared to consonances. Visual perception studies have consistently shown that activity in early visual areas is reduced whenever individual features of an image are perceived as "coherent patterns or shapes" (compared to randomly arranged visual elements) (Murray et al. 2002; Dumoulin and Hess, 2007; Fang et al. 2008). Evidence suggests that these effects could be relevant for inferential processes, contributing to the disambiguation of sensory inputs (Bar et al. 2006). Moreover, it has been proposed that expectations about the precision of sensory inputs may play a central role beyond the dynamics of perception, affecting also higher cognitive functions such as social judgments and theory of mind processes (Lawson, Rees and Friston, 2014). Accordingly, we predicted that reduced activity in areas involved in processing temporary intentions of others (Van Overwalle, 2009; Van Overwalle and Baetens, 2009) would follow the increases in coherence implied by more consonant sound patterns (Frances, 1972; Bharucha and Prior, 1996; Trehub and Unyk, 1991; Schellenberg and Trehub, 1994, 1996a, 1996b). We further anticipated that the less coherent structure and associated negative valence of dissonant sound patterns would demand greater information integration, heightened mental state attribution resources (Young et al. 2010) and modulate brain systems devoted to appraise behaviourally relevant, unexpected and potentially threatening events (e.g. ventral attention network: Corbetta and Shulman, 2002; Corbetta et al, 2008).

EXPLANATORY BOX. Predictive coding, context frames and musical intervals.

Models of predictive coding propose that the human brain continuously generates predictions to estimate the relevant future built on contextual information from the past. These theories suggest that the brain is not merely reactive but also, and essentially, predictive or proactive (Bar, 2007; Enns and Lleras, 2008; Friston, 2005; Friston and Kiebel, 2009; Grossberg, 2009; Mumford, 1992; Rao and Ballard, 1999). According to one of these frameworks the brain relies on memory-based predictions (Bar et al., 2006; Bar, 2007; Bar et al., 2008), which can be generated based on an external sensory input or driven by thought. Importantly, this model emphasizes the role of memory associations as the building block of predictions. Associations are constructed by extracting repeating patterns and statistical regularities from our environment. This "related" information (i.e. objects that tend to be linked on some level) is thought to be clustered in memory structures that have been termed context frames (Bar, 2004; Bar and Ullman, 1996) [earlier described as scripts (Schank, 1975), frames (Minsky, 1974) or schemata (Mandler and Johnson, 1976)], which imply a "global representation of perceptual and semantic associated attributes" (Bar, 2007).

In the present study we focused on the associations elicited by consonant and dissonant sound patterns, and applied them as contextual frames. Evidence has shown that the degree of consonance/dissonance reliably correlates with the valence percept, with consonant musical intervals being consistently associated with more positive valence inferences, compared to dissonant intervals (Blood et al., 1999; Costa et al., 2000; Plomp and Levelt, 1965; Trainor and Heinmiller, 1998). The emphasis in our work is based on the theoretical model first put forward by Rao and Ballard (1999), which proposed that early sensory cortices would only signal the unpredictable components of their input to higher areas, and that predictable stimuli would require less neural activation to be conveyed from lower to higher areas. In agreement, visual perception studies have shown that predictable stimuli are processed with less neural activation at the earliest cortical relays (Alink et al., 2010), such as when individual features of an image are perceived as "coherent patterns or shapes" (compared to randomly arranged visual elements) (Dumoulin and Hess, 2007; Fang et al., 2008; Murray et al., 2002). Within the musical domain, consonant intervals have been proposed to have a "more coherent structure" and facilitated encoding (Frances, 1972; Bharucha and Prior, 1996; Schellenberg and Trehub, 1994, 1996a, 1996b). In our study, we used a quantitative framework to define predictability/coherence in a sonority built with musical intervals. We employed Temperley's Bayesian model of tonalness (2010) as an index of stimuli tonal predictability/coherence. Tonalness has been defined as "the degree to which a sonority evokes the sensation of a single pitched tone" (Parncutt, 1989) in the sense that sonorities with high tonalness evoke a clear perception of a tonal center (Krumhansl, 2001), whilst sonorities with lower tonalness convey more unpredictable and equivocal tonal centres. Temperley (2010) suggested a way to calculate tonalness level, following a Bayesian 'structure–and-surface' approach, as the overall probability of a pitch-class set occurring in a tonal piece. Previous evidence further indicates that the tonalness level of a sonority could represent a quantifiable predictor of emotional valence associations (Bravo, 2014).

Associations can prime not only perception processes, but also higher cognitive functions such as social judgments (Bar, 2007). Mobbs and collaborators (2006) showed that contextual framing can have a direct influence on mental state predictions (Mobbs et al., 2006) by pairing identical faces with either neutral or emotionally salient contextual movies.

Our task required to predict the affective value of a message conveyed via musical intervals. Participants' judgements were considered as inferences of transitory states or temporary inferences (Van Overwalle, 2009; Van Overwalle et al., 2009), which have been found to rely on theory of mind function (Abell et al., 2000; Castelli et al., 2002; Martin and Weisberg, 2003; Saxe and Wexler, 2005; Schultz et al., 2004; Van Overwalle et al., 2009). We hypothesized that distinct consonance/dissonance levels would exert differential influences on mental state attribution substrates during valence inferences. Specifically, we predicted that higher levels of tonalness would be reflected by reduced activity in these areas, as a result of inhibitory feedback projections being sent from contextual memory systems (Bar et al., 2008), and that the less predictable tonal structure and associated negative valence of dissonant sound patterns would demand heightened mental state attribution resources (Young et al., 2010) and modulate brain systems devoted to mark and evaluate motivationally relevant stimuli (Corbetta and Shulman, 2002; Uddin, 2015).

Finally, building on emerging research which suggests that individuals with autism might exhibit difficulties in processing the affective content implied by certain patterns of stimuli (Abell et al. 2000; Castelli et al.

2002; Adolphs et al. 2001; Dalton et al. 2005; Harms et al. 2010), in a third experiment we investigated the relationship between participants' performance on the task and self-reported symptoms of autism, focusing on clinically typical adult participants who varied along a continuum of the autism spectrum (i.e. participants were not selectively chosen for presenting, or not presenting autistic traits). Since autism is a spectrum disorder expressed with vast heterogeneity, both between people diagnosed with autism and within the general population, an advantage of the present approach was that it could enable the assessment of how varying levels of symptoms would relate to performance on the task. Ultimately, the rationale behind the design of the present sound-based paradigm was that, if proven useful, it could be further developed into a non-verbal task applied for assessing valence inferences in groups with specific language impairments.

It is important to note, however, that this behavioural experiment did not test individuals diagnosed with autism, the objective of the study was to investigate the variation in participants' responses to the task in relation to different levels of autistic traits in a non-clinical population.

Materials and Methods

Subjects

Experiment 1 (Laboratory):

Twenty-three individuals participated in the laboratory experiment (11 females, 12 males; mean age = 27.63, SD = 1.70). Subjects reported no long-term hearing impairment. None of the participants was a professional musician, five participants reported having received informal musical training for less than three years. All subjects gave informed consent. The study received ethical approval from the Music Faculty Research Ethics Committee (University of Cambridge), Reference Number 12/13.4.

Experiment 2 (fMRI study):

Data were obtained from sixteen healthy volunteers, of which fourteen were included in the final analysis. Two participants were excluded on the basis missing functional imaging data in the frontal lobes, possibly due to unusual positioning or susceptibility artifacts. The fourteen participants included (7 females, 7 males; mean age = 30.45, SD = 3.05) were right handed, native Spanish speaking born in Argentina (participants were not screened by ethnicity but by first language), from Fundación Científica del Sur Imaging Centre (http://fcsur.com/) community (radiology residents, radiographers and administrative personnel), with no history of neurological or psychiatric illness, or use of psychotropic medication. The study received ethical approval from the Institutional Review Board of Fundación Científica del Sur (Buenos Aires, Argentina).

Experiment 3 (Internet-based):

Thirty-nine individuals (19 females, 20 males; mean age = 31, SD = 6.36) participated in the internet-based experiment. They reported no long-term hearing impairment. None of the participants was a professional musician, one participant reported having received informal musical training for less than four years. All subjects gave informed consent. The study received ethical approval from the Music Faculty Research Ethics Committee (University of Cambridge), Reference Number 12/13.4.

Experiment 3 was designed to investigate the relationship between participants' performance on the SETi task and self-reported symptoms of autism. It focused on adult participants who varied along a continuum of the autism spectrum. It is important to note, however, that this experiment did not test individuals clinically diagnosed with autism.

Stimulus material and design

Auditory stimuli construction:

Tonal dissonance is considered to be one of the most essential aspects that build musical tension (Lerdahl and Krumhansl, 2007). Tonal dissonance has been characterised by three distinctive components: tonal function, melodic organization and sensory dissonance (Bigand et al., 1996; Bigand and Parncutt, 1999). The present experiment was centred on the effect of sensory consonance/dissonance (Helmholtz, 1895). We employed sequential intervals, which do not produce roughness or beats due to their non-simultaneity; however, they are also judged along the dimension of consonance/dissonance according to their frequency ratios (Ayres et al. 1980; Schellenberg and Trehub, 1994). The experimental stimuli for the task were constructed with a stand-alone application (programmed in Max-Cycling'74 by Bravo) that enabled complete randomization and reaction time recordings (laboratory version). The sounds for the experiment were created using pure tones, and systematically manipulated through algorithms, by means of which the three distinct levels of consonance/dissonance were generated. The consonance condition employed a consonant interval content (interval set [5-perfect fourths-, 7-perfect fifths-, 12-octaves-]), the intermediate dissonance condition was built based on the diminished triad (interval set [3-minor thirds]); finally, the strong dissonance condition was constructed with a dissonant interval content (interval set [1-minor seconds-, 2-major seconds-, 6-tritones-]). Table 1 shows the three pitch-class sets that were employed in this experiment (corresponding to the three sound conditions described above) together with their respective *tonalness* values. The term tonalness has been defined as "the degree to which a sonority evokes the sensation of a single pitched tone" (Parncutt, 1989) in the sense that sonorities with high tonalness evoke a clear perception of tonal pitch center. As a component of consonance, it has been characterised as the ease with which the ear/brain system can resolve the fundamental, being the easier, the more consonant. Temperley (2007) has suggested a way to calculate tonalness level, following a Bayesian 'structure–and-surface' approach, as the overall probability of a pitch-class set occurring in a tonal piece. Empirical evidence indicates that tonalness could represent a quantifiable predictor of emotional valence associations (Bravo, 2014).

Table 1. Tonalness values for the three sound conditions utilised in the experiment, calculated using the Kostka-Payne key-profiles (Kostka, 2003; Kostka et al., 2012) (Numbers in brackets [ ] indicate interval set, numbers in parenthesis ( ) denote the correspondent prime form.

Interval Set/ Prime Form	Tonalness
Consonance [5,7,12]/(0,2,7)	0.00231
Intermediate dissonance [3]/(0,3,6)	0.00032
Strong dissonance [1,2,6]/(0,1,2,6)	0.00016

The impact of each sound category on the three affective dimensions of valence, arousal and potency (Bradley & Lang, 1994; Lang, Greenwald, Bradley, & Hamm, 1993; Russell, 1979) had been preliminary validated in a pilot study with 135 naive normal subjects (age range: 26-30), tested during the 2012 Cambridge's Festival of Ideas. An ANOVA test, conducted to assess whether there were differences between the ratings (i.e. judgements of valence, potency and arousal) for each sound condition, revealed no significant differences either in the arousal dimension ($F_{2, 102} = 1.68$, $P = 0.190$), or in the potency dimension ($F_{2, 102} = 0.147$, $P = 0.863$). Results showed that participants did rate the sound conditions differently in the valence dimension ($F_{2, 102} = 3.84$, $P = 0.025$). Post hoc contrasts (Tukey HSD tests, equal variances assumed) indicated that the valence ratings for the strong dissonance and the consonance conditions differed significantly ($P = 0.019$, 95% CI [-1.86, -0.13]).

Supported by previous evidence (reviewed in: Schellenberg and Trehub, 1994), an assumption was made that participants' valence judgement would be influenced by the level of consonance/dissonance, with consonant sounds leading to positive interpretations of the auditory signals, whilst increasing levels of dissonance would guide participants towards more ambiguous and negatively valenced evaluations.

The complete experiment (Figure 1, left) involved 24 blocks of sound. Each six-second block of sound consisted of 41 musical interval presentations (42 individual notes) for a particular condition. Although only one block of sound per category was found to be sufficient to elicit distinctive valence inferences between conditions at a behavioural level [pre-test with n = 26; significant differences observed between strong dissonant and consonant sounds; F (1, 25) = 13.632, $p < 0.001$], we designed the experiment to include 8 blocks of sound per condition (totalling 328 musical interval presentations per sound category), in order to reliably estimate the haemodynamic response function (HRF) and to show detectable differences between conditions in the neuroscientific setting. Each six-second block of sound started with a distinct, randomly assigned, initial pitch (i.e. each sound block was unique), but which belonged to the intervallic-content set determined by each sound condition. Four different pseudo-randomized orderings of the sound blocks were utilised, in which sound blocks were carefully distributed to avoid contrasting trials that are far apart in time in the fMRI analysis. The session always started with a sound block. A silent condition was added with 8 presentations of six seconds each (blocks of rest). We utilised the blocks of rest as basic baseline to identify sound encoding during task performance. Sound blocks were separated by two seconds of silence (inter-trial interval), unless there was a silent condition in between two sound blocks, in which case no additional separation time was included. No repetitions of silence were allowed, and there were never more than two consecutive sound blocks belonging to the same level of consonance/dissonance. Sound parameters were set as follows (identical for all three sound conditions): within a sound block, each note had a total duration of 128 milliseconds (ms), including 10-ms raised-cosine onset and offset ramps, and was triggered with a fixed velocity (i.e. constant loudness). Notes were separated by 15-ms gaps, producing an overall presentation rate of 7 notes per second (42 notes = 41 musical intervals per six-second sound block).

Design:

A repeated measures design was employed. The manipulated (independent) variable was the level of consonance/dissonance (consonance, intermediate dissonance and strong dissonance). The outcome (dependent) variable was participants' ratings in terms of valence inferences (positive or negative). This was measured using categorical binary responses (laboratory experiment) and 5-point Likert scales (internet-based experiment). Behavioural data was obtained in the laboratory experiment through a dichotomous format since in the fMRI experiment participants were required to listen to the sounds, and to respond covertly ("...to think and decide if they were produced by good-friendly or bad-aggressive aliens"), and in this context, the answer to the task was assumed to correspond more to a binary response format than a to a multiple-categories rating format. Detailed ratings (i.e. 5-point scales with the extremes labelled 'good-friendly aliens' and 'bad- aggressive aliens', order counterbalanced) were employed in internet-based experiment to obtain higher resolution/more fine-grained scores to perform statistical analyses with. In the internet-based experiment, the level of accuracy in valence ratings was calculated by counting the number of correct responses ascribed to the consonant sounds (where correct answers = ratings > 3). Since there were eight presentations for each condition, the variable accuracy level could vary between 0 and 8, with 0 representing no correct response and 8 complete accuracy. Familiarity ratings were included in the fMRI experiment (participants were presented with excerpts from each of the three sound categories in pseudo-randomized order, and asked to rate their familiarity on a five-point bipolar scale ranging from 1-unfamiliar to 5-familiar).

Procedure

The same task, using exactly the same stimuli, was carried out in all three experimental settings (i.e. laboratory, fMRI and internet-based). Subjects were asked to make temporary valence inferences about imaginary others based on non-verbal auditory cues. The task (described in Figure 1) utilised a purposely-made metaphor, which informed the participants that a radio-telescope located in Cambridge had captured a series of radio signals from outer space. Participants were required to listen to these radio signals, and to "decide if they were produced by good-friendly or bad-aggressive aliens". In the present study, participants' judgements were inferences of transitory states or temporary inferences (Van Overwalle, 2009), which have been found to rely on theory of mind function (Castelli et al. 2002; Martin and Weisberg, 2003; Schultz et al., 2004; Saxe and Wexler, 2005; Van Overwalle et al., 2009).

Experiment 1 (Laboratory): The experiment was run in the Centre for Music and Science (CMS) at Cambridge University. All subjects performed the task using the CMS workstations and listened to the stimuli with Behringer HPM1000 Headphones. Sound pressure levels were measured with a Galaxy Audio CM130 Meter, the output volume was set to be identical in all workstations (average sound level = 70 dB). Participants had to select their answer for each sound by clicking the correspondent 'good-friendly' or 'bad- aggressive' alien image (side semi-randomized). The laboratory version, a stand-alone application (programmed in Max-Cycling'74 by Bravo), permitted reaction time recordings (captured in milliseconds). Observed test performance in a controlled setting was essential to validate the performance of subjects in the internet-based sound experiment.

Experiment 2 (fMRI study):

Participants were asked to arrive to the Imaging Centre 45 minutes before the fMRI scanning session, in order to undertake the training session of 10 minutes in a separate room (contiguous to the scanner room). Subjects were familiarised with the task and trained on the procedure with nine trials (three per sound condition) with sample stimuli constructed based on the testing materials. Participants were instructed to think and decide on a response to the task question as soon as they heard the onset of each of the sound blocks, which were separated by blocks of silence.

In the fMRI setup, the visual stimulus (invariant still image of a radio-telescope) was projected onto a screen and presented to the subject via a 45° angled mirror positioned above the participant's head. The auditory stimuli were delivered via Etymotic ER30 tube-phones (Etymotic Research, Illinois, USA). Subjects were given the same instruction as in the behavioural study, but they were asked to produce a covert response. Following the scanning session each subject underwent the internet-based version of the experiment (using the same order for the sound conditions as inside the MRI scanner) in order to collect subject specific behavioural data, which was subsequently related to the functional imaging data.

Experiment 3 (Internet-based):

Participants underwent the internet-based experiment individually. The internet-based experiment contained two sections. The first section involved the sound experiment (SETi). In the second section, participants were assessed with the Autism Spectrum Quotient questionnaire (AQ: Baron-Cohen et al. 2001), a self- administered personality questionnaire that is normally used as a dimensional measure of autistic traits. The AQ test aims to identify the degree to which any individual adult of normal intelligence might have features of the core autistic phenotype (i.e. it investigates whether adults of average intelligence have symptoms of autism spectrum conditions). The fifty questions on the AQ questionnaire are made up of ten questions assessing five dimensions: imagination, communication, attention to detail, attention switching and social skills. The 50-item self-report AQ questionnaire has been shown to have good discriminative validity and good screening properties (Woodbury-Smith et al. 2005). The AQ questionnaire aims to identify the degree to which any individual adult of normal intelligence might have features of the core autistic phenotype (i.e. it investigates whether adults of average intelligence have symptoms of autism spectrum conditions) (Baron-Cohen et al., 2001). The fifty questions on the AQ questionnaire are made up of ten questions assessing five different subcategories or dimensions: imagination, communication, attention to detail, attention switching and social skills. The 50-item self-report AQ questionnaire has been shown to have good discriminative validity (i.e. it differentiates between patients who received a diagnosis of Asperger Syndrome according to the DSM-IV diagnostic criteria and those who did not) and good screening properties at a threshold score of 26 (Woodbury-Smith et al., 2005). In the preliminary study conducted by Woodbury-Smith and collaborators (2005) a threshold score of 26 resulted in the correct classification of individuals with AS (at this cut off the sensitivity was 0.95, specificity 0.52, positive predictive value 0.84, and negative predictive value 0.78). The authors proposed that a threshold score of 26 would ensure that false negatives are limited, while equally avoiding false positives (Woodbury-Smith et al., 2005). They suggested that a higher cut off score of 32 should be utilised in general population screens (Baron-Cohen et al., 2001; Woodbury-Smith et al., 2005), and further advised that within the general population “there may be a percentage of individuals who have many autistic traits but who do not require any clinical support (and are not seeking this) because of a good cognitive match between their cognitive style or personality, and their family or occupational or social context” (Woodbury-Smith et al., 2005, p. 334) (BaronCohen, 2003). Accordingly, whether a high AQ score becomes disabling “may depend on environmental factors (being valued for contribution at work, tolerance by significant others, or a place in a social network protecting against the risks of secondary depression) rather than solely on factors within the individual” (Woodbury-Smith et al., 2005, p. 334). Total scores can range from 0 to 50, with higher scores implying more symptoms of autism spectrum disorders (ASD). The a priori hypothesis was that the AQ score would correlate with participants’ performance in the sound experiment. Specifically, we hypothesized that either a) participants with more autistic traits would accurately rate all three sound categories (e.g. consonant sounds would be appraised as positively valenced, strong dissonant sounds as negatively valenced, and intermediate dissonant sounds would be rated in between them) or, b) participants with more autistic traits would attribute inappropriate valence to all sound conditions. Both hypotheses were based on Baron-Cohen’s Empathizing-Systemizing (E-S) theory (Baron-Cohen et al. 2001). The first hypothesis was based on E-S theory prediction of normal (in line with mental age) or superior systemizing if a domain is systemizable (and music is a systemizable domain according to the E-S theory). The second hypothesis was based on E-S theory prediction of impaired empathizing. The present study entails a mental state attribution task and, therefore, abnormal empathizing capabilities could impact on this specific function.

fMRI data acquisition

A General Electric Signa system operating at 3 Tesla was used. Prior to the functional magnetic resonance measurements, high resolution (1 x 1 x 1 mm) T1-weighted anatomical images were acquired from each participant using three-dimensional fast spoiled gradient- echo (3D-FSPGR) sequence. Continuous Echo Planar Imaging (EPI) with blood oxygenation level-dependent (BOLD) contrast was utilised with a TE of 40ms and a TR of 3000ms. The matrix acquired was 64 x 64 voxels (in plane resolution of 3 mm x 3 mm). Slice thickness was 4 mm with an interslice gap of 0.7 mm (35 slices with interleaved acquisition, whole brain coverage). Functional images were acquired over one run of 4 minutes. The sound files used for the task were digitally recorded onto compact disks and delivered to participants at a loudness level equal for all subjects.

fMRI Data analysis

Data were processed using Statistical Parametric Mapping (SPM), version 8 (Wellcome Department of Imaging Neuroscience, London, UK). Following correction for the temporal difference in acquisition between slices, EPI volumes were realigned and resliced to correct within subject movement. A mean EPI volume was obtained during realignment and the structural MRI was coregistered with that mean volume. The coregistered structural scan was normalized to the Montreal Neurological Institute (MNI) T1 template (Friston et al. 1995). The same deformation parameters obtained from the structural image, were applied to the realigned EPI volumes, which were resampled into MNI-space with isotropic voxels of 3 cubic millimeters. The normalized images were smoothed using a 3D Gaussian kernel and a filter size of 6 mm FWHM. A temporal highpass filter with a cutoff frequency of 192 Hz was applied with the purpose of removing scanner attributable low frequency drifts in the fMRI time series. An event-related design was modeled using a canonical hemodynamic response function and its temporal derivative. The design matrix included the following four regressors: consonant sounds, strong dissonant sounds, intermediate dissonant sounds and rest (baseline). Parameter estimate images were generated. Nine contrast images per individual were calculated: strong diss > cons, intermediate diss > cons, strong diss > intermediate diss, intermediate diss > rest, strong diss > rest, cons > rest, cons > intermediate diss, cons > strong diss and intermediate diss > strong diss. Second level group analyses were carried out using one-sample t-tests. The significant map for the group random effects analysis was thresholded at voxel level $p < 0.001$ uncorrected, with a cluster level threshold of $p < 0.05$ corrected using FWE (family wise error). Whole-brain analyses were performed for the linear contrasts that compared sound conditions against the baseline condition. The analysis of all the linear contrasts that comprised sound conditions between each other was restricted to regions of interest that were defined based on a) meta-analytic reviews (statistical summaries of empirical findings across studies) of ToM and ventral attention network regions, and b) previous neuroscientific studies that have investigated the neural response to tasks involving the emotional evaluation of musical dissonance (Table 2).

Table 2. Regions of interest for fMRI analysis of linear contrasts that comprised sound conditions.

Region of Interest	Motivation	References
Parahippocampal Gyrus	Neural correlates of emotional responses to dissonance	Blood et al. 1999; Koelsch et al. 2006; Gosselin et al. 2006
Anterior cingulate cortex and medial prefrontal cortex	Neural correlates of emotional responses to dissonance/ Evaluation of emotional content in music and voice/ Mental state attribution	Green et al., 2008; Foss et al., 2007/ Escoffier et al., 2013/ Saxe and Kanwisher, 2003; Saxe and Wexler, 2005
Right Angular Gyrus	Processes requiring intentional attribution	Saxe, 2006, 2010, Scholz et al. 2009; Carter and Huettel, 2013;

	and temporary inferences	Van Overwalle, 2009
Bilateral insula	Ventral attention and salience networks (marking behaviourally relevant events for additional processing)	Corbetta and Shulman, 2002; Corbetta et al, 2008; Menom and Uddin, 2010; Uddin, 2015

Meta-analytic reviews have provided evidence supporting the role of midline cortical structures (medial prefrontal, anterior and posterior cingulate cortices) as well as bilateral temporo-parietal junction areas in mental state attribution (Saxe and Kanwisher, 2003; Saxe and Wexler, 2005). With respect to the present study, activation in the right angular gyrus seems to be strongly linked to processes requiring intentional attribution and temporary inferences (rTPJ-Mental: Saxe, 2006, 2010, Scholz et al. 2009; Carter and Huettel, 2013; Van Overwalle, 2009). Core regions of the ventral attention network include the rTPJ, ventral frontal cortex and bilateral anterior insula (Corbetta and Shulman, 2002; Corbetta et al, 2008). Small volume correction was also applied to signal changes observed in the parahippocampal cortex (PHC) since convergent evidence has supported its function in the emotional appraisal of musical dissonance (Blood et al. 1999; Koelsch et al. 2006; Gosselin et al. 2006). All ROIs were defined using anatomical masks of the described areas with WFU PickAtlas Toolbox (Maldjian et al. 2003).

Psycho-physiological interactions (PPI) analysis: Following the approach developed by Friston et al. (1997) functional connectivity was measured in terms of psycho-physiological interactions (PPI). Seed regions of interest (in the left parahippocampal cortex) were selected on the basis of significantly activated clusters from the subtractive analysis comparing consonance, intermediate dissonance and strong dissonance against the baseline condition. The group cluster peaks (e.g. consonance > baseline: MNI coordinates -18 -28 -17) were used as point of reference to identify individual subject activation peaks that complied with the following two rules: a) were within a 24 mm radius, and b) were within the boundaries of the corresponding brain area (created using the WFU pickatlas toolbox: Maldjian et al. 2003). After the identification of the relevant statistical peaks for each subject, spheres were defined around these peaks with a 6mm radius, which were used as the seed regions of interest for the Psychophysiological Interaction (PPI) analysis. This type of analysis is used to detect target regions for which the covariation of activity between seed and target regions is significantly different between the experimental conditions of interest: consonance > baseline, intermediate dissonance > baseline and strong dissonance > baseline. For each seed ROI, the contrast images from all subjects were used in voxel-wise one-sample *t*-tests at the second level (at threshold level $p < 0.001$ voxel uncorrected, $p < 0.05$ cluster FWE-corrected).

Dynamic Causal Modeling (DCM) analysis (effective connectivity): DCM utilises the temporal information contained in fMRI data to estimate and make inferences about the causal relationships of activity patterns between different brain areas (Friston et al. 2003). In the present study, DCM was carried out to investigate the interaction between the left parahippocampal cortex (lPHC) and the right angular gyrus (rAG). We predicted that distinct consonance/dissonance levels would exert differential influences on mental state attribution substrates during valence inferences. We anticipated that reduced activity in areas involved in processing temporary intentions of others (rAG: Van Overwalle, 2009; Van Overwalle and Baetens, 2009) would follow the increases in coherence implied by more consonant sound patterns. Based on predictive coding models (Rao and Ballard 1999, Friston, 2009) we hypothesized that this effect would result from inhibitory feedback projections being sent from contextual memory systems (i.e. parahippocampal cortex - PHC). To test this hypothesis we carried out an effective connectivity analysis using DCM (Friston, 2003).

DCM can be applied to networks with as few as two nodes, and some of its most powerful applications have been performed with very simple models (Stephan et al. 2010). In this study DCM analysis was conducted for each subject on two ROIs: the lPHC and the rAG. Coordinates for defined areas are detailed in Table 3. The modulatory effect of interest was the consonant condition, since this category appeared to strongly modulate the functional interaction between these two areas (as shown in the PPI analysis). It is important to note that there is no circularity in using the same data to define regions by SPM and analyse their interactions with DCM. This is because, in contrast to SPM, the purpose of DCM is not to test whether any of these regions show an experimental effect but to compare different hypotheses about directional transfer of information mechanisms (in terms of neuronal coupling) that underlie the regional responses detected in the analyses.

Table 3. Regions of interest for dynamic causal modeling (DCM) analysis, describing peak activation MNI coordinates for clusters examined, mask definition and contrasts for VOI extraction (PHC: parahippocampal cortex; AG: angular gyrus).

Region of Interest	Cluster peak	Contrast for Mask	Contrast for VOI extraction	Modulatory effects examined
Left Parahippocampal	-18 -28 -17	Consonance > Baseline	Cons>Strong Diss.	Consonance
Right Angular Gyrus	30 -55 43	PPI analysis seed in left PHC (Cons>Baseline)	Cons>Strong Diss.	Consonance

For each subject, three models were defined. Model 1 specified a connection from the lPHC to the rAG (hypothesized). Model 2 specified a connection from the rAG to the lPHC. Model 3 specified bidirectional connections between both regions. To explore how consonance induced changes in connectivity between brain areas, this condition was included as a modulatory effect allowing it to change any connection in the model. A random effects approach was used since it could not be assumed that the best fitting model structure would be constant across subjects (Stephan et al. 2010). Bayesian model selection (BMS) was utilised to select the optimal model (Stephan et al. 2010). DCM estimates three kinds of coupling parameters for a given model: (i) Direct influences of driving inputs on the neuronal states, (ii) strengths of intrinsic connections that reflect the context-independent coupling between neuronal states in different regions, and (iii) modulatory or bilinear inputs that reflect context-dependent changes (i.e. produced by the experimental conditions) in the coupling between regions. Extrinsic parameters were not estimated in the present analysis, since the process of interest specifically targeted modulatory parameters, which measured changes in effective connectivity induced by consonance. The parameter estimates describe the speed at which the neural population response changes, which has an exponential decay nature (Stephan et al. 2010). Therefore, parameters are expressed in terms of the rate of change (unit: Hz) of neuronal activity in one area that is associated with activity in another, and can be either positive or negative. A positive parameter means that an increase in activity in one region results in increased rate of change in the activity of another region. Conversely, a negative parameter means that an increase in activity in one region results in a decreased rate of change in the activity of another region. Dynamic causal modeling was performed using DCM10 in SPM8 software (http://www.fil.ion.ucl.ac.uk/spm).

Results

Experiment 1 (Laboratory)

Twenty-three participants were asked to respond to the SETi task in a controlled experimental setting. Using a multivariate approach, results indicated that participants did rate the three sound conditions differently (Wilks' Lambda $F_{2,\ 21}$ = 42.853, P < 0.001). Post hoc contrasts (Bonferroni corrected) showed that the average valence rating for consonant sounds was significantly more positive than the valence rating for strong dissonant (P < 0.001, d = 2.565) and intermediate dissonant sounds (P < 0.001, d = 2.391). There was no significant difference in valence ratings between strong and intermediate dissonant sounds (P > 0.999). Polynomial contrast on the mean ratings for the three sound categories (listed in Table 4) indicated a significant linear trend ($F_{1,\ 38}$ = 6.276, P < 0.01), confirming that participants gave more extreme valence ratings to stimuli with more extreme consonant (or dissonant) interval content, whilst intermediate dissonances were evaluated as moderate in valence.

A repeated measures ANOVA, with Greenhouse-Geisser correction (sphericity assumption violated), assessing whether there were differences between the reaction times (RTs) for the three sound conditions yielded significant differences ($F_{1.08,\ 23.79}$ = 4.27, P = 0.047). Participants took considerably longer to rate the intermediate dissonant sounds than the other two conditions. The difference was significant between intermediate dissonant and consonant sounds (pairwise comparisons, $F_{1,\ 22}$ = 4.74, P = 0.04), and close to significance for intermediate dissonant vs. strong dissonant sounds ($F_{1,\ 22}$ = 3.98, P = 0.058). We found no significant differences in RTs between the strong dissonance and consonance conditions (Table 4).

Table 4. Valence means with standard deviations (proportion times [out of 8] that each sound was categorized as 'good', higher numbers denote more positive valence), and reaction times means with standard deviations (shown in milliseconds) for each sound condition.

Composite rating	Valence Mean	(s.d.)	R. Time Mean	(s.d.)
Consonance	3.4347	(0.843)	2405.43	(2357.72)
Intermediate Dissonance	1.0434	(1.021)	7010.93	(9529.17)
Strong Dissonance	0.8695	(0.868)	2866.30	(1945.69)

Overall, participant's evaluation of consonances (which were distinctively linked to positive valence) conveyed the fastest reaction times; however, they were only significantly faster than the valence judgments for intermediate dissonances, which elicited the slowest reaction times.

Experiment 2 (functional Magnetic Resonance Imaging study).

Participants in the fMRI study were given the same instruction as in the laboratory setting, but were asked to produce a covert response (i.e. "...listen to these sounds and your task is to *think and decide* if they were produced by good-friendly or bad-aggressive aliens..."). Following the scanning session each subject underwent the internet-based version of the experiment (same ordering of sound stimuli as inside the MRI scanner) to collect subject-specific behavioural data, which was subsequently related to the functional imaging data. No significant differences were found in the familiarity ratings of the three sound categories ($F_{2,12}$ = 0.172, P = 0.844), which appeared to be similarly unfamiliar to all subjects. Results indicated that participants rated the conditions differently (Wilks' Lambda $F_{2,12}$ = 6.91, P = 0.008). Post hoc tests (Bonferroni corrected) showed that the valence rating for consonant sounds was on average significantly more positive than the valence rating for strong dissonant (P = 0.005, d = 0.491), supporting the findings reported for Experiment 1. Polynomial contrasts revealed the same significant linear trend for valence ratings ($F_{1,13}$ = 15.525, P = 0.002), strong dissonance < intermediate dissonance < consonance, although the intermediate dissonant condition did not differ significantly from either of the other two conditions.

A markedly similar pattern of brain activation was found when contrasting each sound condition against the baseline condition, comprising clusters in bilateral superior temporal regions and in the left parahippocampal cortex (Table 5A, Figure 2a, Figure S5). Statistical peaks in Heschl's gyri (primary auditory cortex) corroborated basic responses to auditory stimuli. The observation of neural responses in the left parahippocampal cortex (lPHC), an area implicated in contextual memory (Aminoff et al. 2013), is consistent with previous literature reporting parahippocampal activity in response to processing emotional aspects of music, and specifically, in response to the emotional evaluation of musical stimuli with contrasting levels of dissonance (Blood et al. 1999; Koelsch et al. 2006; Gosselin et al. 2006).

The comparison between consonant and intermediate or strong dissonant sounds did not reveal any suprathreshold signal changes (Table 5A). Whilst participants were making valence inferences ascribed to the strong dissonant sounds (compared to consonant sounds) activation was observed within a cluster comprising the right angular gyrus (rAG) and the right inferior parietal cortex (Table 5A, Figure 2b). In agreement with previous research revealing increased ToM processing for negatively valenced judgments that rely on mental state information (Young et al. 2010), higher levels of dissonance induced a stronger response in the rAG, an area specifically implicated in ToM function (Saxe and Wexler, 2005; Saxe, 2010). The contrast between strong dissonance and intermediate dissonance revealed significant activations in the right anterior cingulate cortex (ACC) and in the anterior insula (AI) bilaterally (Table 5A). Emerging evidence indicates that bilateral AI and ACC form a “salience network” (Menom and Uddin, 2010) that is sensitive to salient and environmental stimuli, and which core function is to mark and segregate such events in time and space for additional processing.

Table 5. (A) Results (FWE-corrected $P < 0.05$ for cluster-level inference) of group General Linear Model for the contrasts: consonance > baseline, strong dissonance > baseline, intermediate dissonance > baseline, strong dissonance > consonance, consonance > intermediate dissonance and consonance > strong dissonance. (B) Results of group PPI analysis for the contrasts: consonance > baseline, intermediate dissonance > baseline and strong dissonance > baseline, with seed voxels located in the left parahippocampal cortex (PHC). The regions described showed stronger positive functional connectivity with the PHC. (C) Dynamic Causal Modeling (DCM) Bayesian model selection (BMS) results: Conditional probability (expected posterior probability representing the probability of a model given the observed data) and exceedance probability (probability compared with other tested models). Model 1 specified a connection from the left parahippocampal cortex to the right angular gyrus, model 2 specified a connection from the right angular to the left parahippocampal cortex, and model 3 specified bidirectional connections. Model 1, in which the right angular gyrus received information from the left parahippocampal cortex obtained the most evidence (99%). Abbreviations: L: left, R: right.

Region	Peak MNI	Voxels	t (z-value) [voxel $p < 0.001$]	Region Mean t (std.)	p-value [Cluster FWE-corrected]
A (Subtractive analysis)					
Consonance > Baseline					
Temporal Sup R	45 -16 4	356	11.01 (5.42)	5.56 (1.22)	< 0.001
Heschl R		48		6.58 (1.73)	
Temporal Sup L	-54-13-1	293	9.38 (5.08)	5.54 (1.13)	< 0.001
Heschl L		36		5.07 (0.92)	
Parahipp. L	-18 -28 -17	8	5.71 (3.97)	4.93 (0.43)	0.023
Intermediate Dissonance > Baseline					
Temporal Sup R	51 -7 -2	351	7.91 (4.49)	5.29 (0.98)	< 0.001
Heschl R		53		6.26 (1.34)	
Temporal Sup L	-54 -16 -2	278	9.19 (5.04)	5.39 (1.08)	< 0.001
Heschl L		44		4.99 (0.90)	
Parahipp. L	-15 -28 -11	5	5.53 (3.90)	4.59 (0.70)	0.034
Strong Dissonance > Baseline					
Temporal Sup R	51 -7 -2	346	10.02 (5.22)	5.37 (1.09)	< 0.001
Heschl R		50		6.13 (1.35)	
Temporal Sup L	-54 -16 -2	272	11.25 (5.47)	5.42 (1.30)	< 0.001
Heschl L		35		5.12 (0.95)	
Parahipp. L	-15 -28 -14	3	4.64 (3.50)	4.63 (0.71)	0.047
Strong Dissonance > Consonance					
Angular R	36 -58 43	26	6.18 (4.15)	4.56 (0.67)	0.013
Parietal inferior R	51 -40 52	7	5.20 (3.76)	4.36 (0.41)	0.032
Strong Dissonance > Intermediate Dissonance					
Insula L	-30 23 -8	2	4.47 (3.42)	4.17 (0.42)	0.005
Insula R	45 14 -8	3	3.84 (3.09)	3.75 (0.07)	0.022
Anterior Cingulate R	9 41 19	11	4.57 (3.47)	3.53 (0.45)	0.021
Consonance > Intermediate dissonance				No suprathreshold signal changes	
Consonance > Strong dissonance				No suprathreshold signal changes	

B (Functional connectivity analysis)					
PPI - Consonance > Baseline (Seed 6mm sphere in left PHC, MNI -18 -28 -17)					
Angular R	30 -55 43	9	6.48 (4.26)	4.81 (0.79)	0.029
Parietal inferior R		5		4.45 (0.38)	
PPI – Intermediate dissonance > Baseline (Seed 6mm sphere lPHC, MNI -15 -28 -11)				No suprathreshold signal changes	
PPI – Strong dissonance > Baseline (Seed 6mm sphere in lPHC, MNI -15 -28 -14)				No suprathreshold signal changes	
C (Effective connectivity analysis)					
DCM – BMS (Effects of consonance) Models			Expected $< sk \mid Y >$	Exceedance ψk	
1) Left PHC → right AG			0.8343	0.9972	
2) Right AG → left PHC			0.1043	0.0025	
3) Bidirectional			0.0614	0.0003	

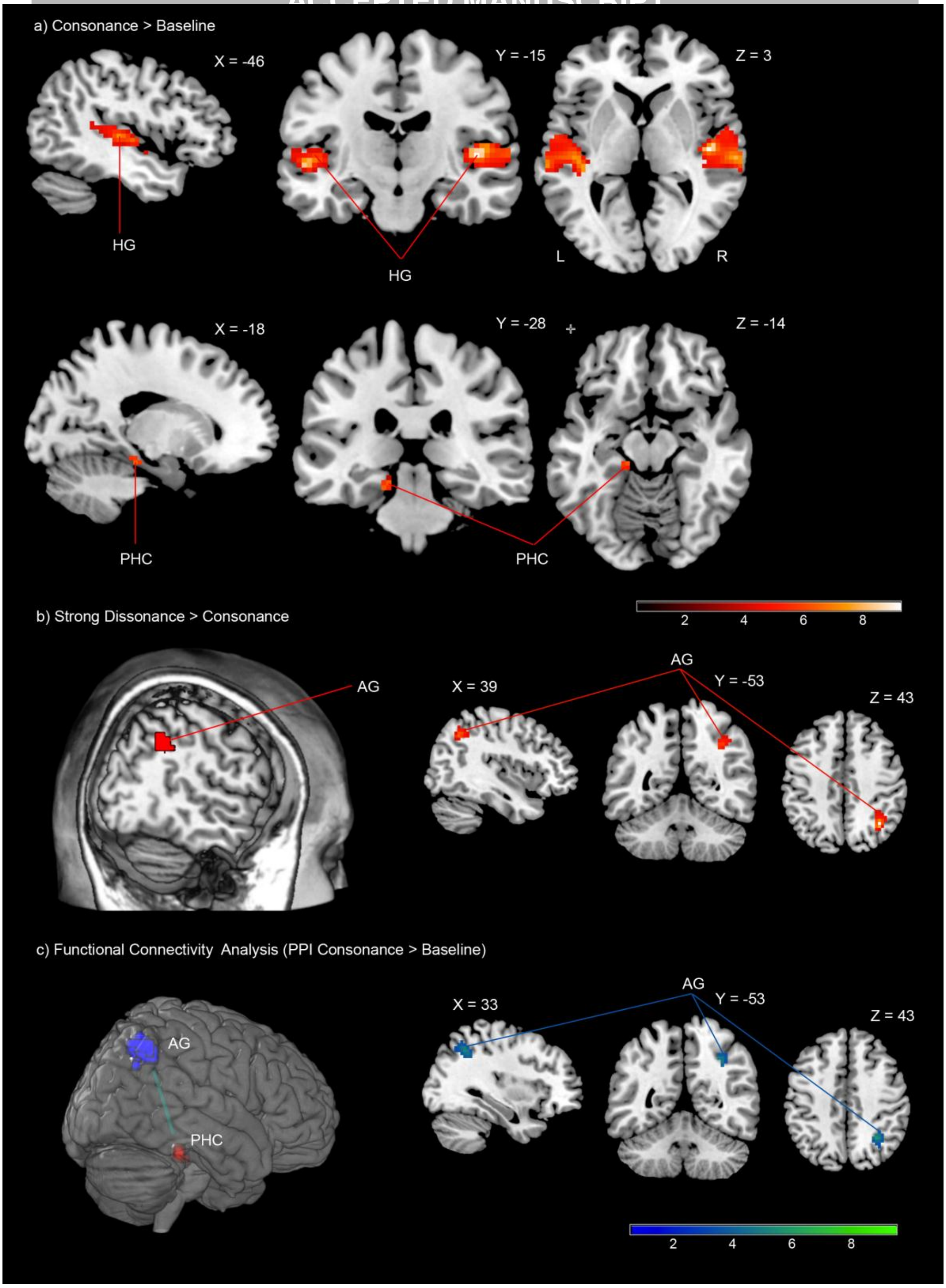


Fig. 2. FMRI results (FWE-corrected $P < 0.05$ for cluster-level inference). Coloured areas (red) reflect: (a) Statistical parametric maps (SPM) of the direct contrast between consonance and rest superimposed onto a standard brain in stereotactic MNI space. Stronger BOLD signals during consonance (compared to baseline) were yielded in temporal superior regions, including Heschl's gyri (HG) bilaterally, and in the left parahippocampal cortex (PHC) (sagittal, coronal and axial views). (b) Statistical parametric maps showing voxels in the right angular gyrus (AG) in which the response was higher during the evaluation of strong dissonant sounds compared to consonant sounds (from left to right: 3D rendering sagittal oblique, sagittal, coronal and axial views). (c) Blue colour identifies voxels in the right angular gyrus, which exhibited stronger functional connectivity with seed voxels (6mm sphere) located in the left parahippocampal cortex during the evaluation of consonant sound processing compared to the baseline condition.

Functional and effective connectivity analyses.

To assess our hypothesis concerning inhibitory feedback projections being sent from memory to theory of mind substrates while participants evaluated consonant patterns, psycho-physiological interaction (PPI) analysis was first performed for the contrast consonance > baseline, with a seed region defined as a sphere with a 6mm radius around MNI coordinates -18 -28 -17 (group cluster peak activation in lPHC for the contrast consonance > baseline, which was used as point of reference to identify individual subject activation peaks). A cluster comprising the rAG and right inferior parietal cortex exhibited stronger positive functional connectivity with the lPHC, supporting a modulatory effect of consonant sounds in the interaction between these two brain structures (Table 5B, Figure 2c). PPI analyses were also conducted for the contrasts intermediate dissonance > baseline, and strong dissonance > baseline (MNI coordinates -15 -28 -11 and -15 -28 -14, group cluster peak activations in lPHC for the respective contrasts); however, no regions survived the defined statistical threshold for cluster-level inference.

To further examine the causal flow of information between the lPHC and the rAG during the evaluation of consonant sound patterns, dynamic causal modeling (DCM) analysis was conducted. A model in which the rAG received information from the lPHC (Model 1) obtained the most evidence (Table 5C). Moreover, Bayesian model averaging demonstrated a negative modulation effect of consonance (-0.0013) on the connection from the lPHC to the rAG that changed the strength of the intrinsic connection (0.4456). This inhibitory effect supports our hypothesis and may account, at least partially, for the decrease in activity observed in the rAG when comparing the evaluation of consonant against strong dissonant sound patterns. The results are consistent with a role of inhibitory feedback signals being sent from memory to mental state attribution substrates, enabling a more efficient/rapid processing of the auditory cues; and appear to parallel, in the theory of mind domain, the evidence from previous studies which have shown deactivations in lower structures with increases in coherence (Murray et al. 2002; Fang et al. 2008).

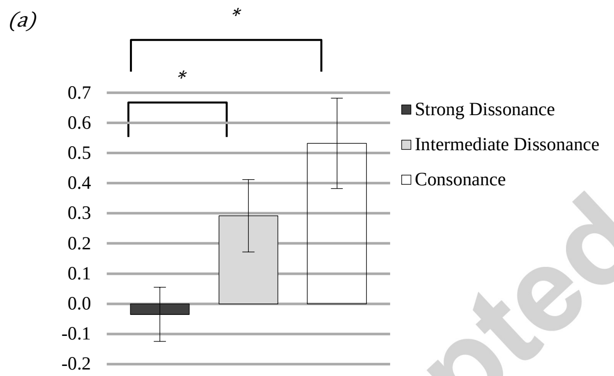
Experiment 3 (Internet-based): Relationship between autistic traits and performance on the task

Relationship between valence ratings and Total AQ score.

Associations between autism spectrum traits and impairments in emotional processing have been observed; however, studies directly assessing emotion recognition in musical structures have, so far, failed to find impairment in this domain (Heaton et al. 1999; Allen et al. 2013). The purpose of experiment 3 was to assess whether the present paradigm could be applied to evaluate emotion processing characteristics in clinically typical adult participants (n = 39) with varying levels of autism spectrum traits.

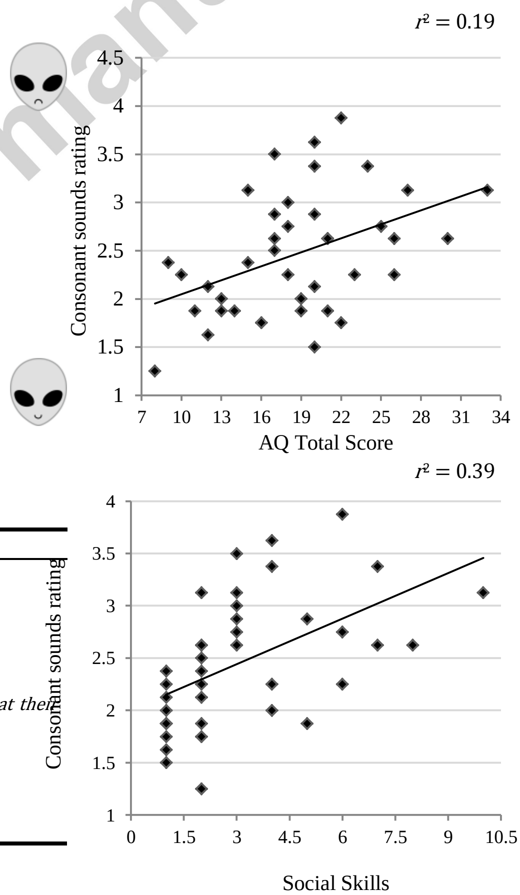
A repeated measures MANOVA indicated that participants rated the three sound conditions differently (Wilks' Lambda $F_{2, 37} = 19.657$, $P < 0.001$). Post hoc pairwise comparisons (Bonferroni corrected) indicated that the valence rating for strong dissonant sounds was on average significantly more negative than the valence rating for consonant sounds ($P < 0.01$, d = 0.567) and intermediate dissonance sounds ($P < 0.01$, d = 0.327). Valence ratings for intermediate dissonant and consonant sounds did not differ significantly (Figure 3a). These results are consistent with the significant differences between strong dissonant and consonant sounds observed in Experiments 1 and 2.

(a)



Composite rating	Mean	Std. deviation	95% CI (adjusted)
Strong Dissonance	3.035	0.712	[2.933, 3.123]
Interm Dissonance	2.708	0.654	[2.585, 2.817]
Consonance	2.467	0.628	[2.315, 2.606]

(b)



(c)

Social Skills AQ Subscale Items
I prefer to do things with others rather than on my own.
I find social situations easy.
I would rather go to a library than a party.
I find myself drawn more strongly to people than to things.
I find it hard to make new friends.
I find it easy to work out what someone is thinking or feeling just by looking at their face.
I enjoy social occasions.
I find it difficult to work out people's intentions.
I enjoy meeting new people.
I am a good diplomat.

Fig. 3. (a) Composite-valence ratings (y-axis) for strong dissonant, intermediate dissonant and consonant sounds (x-axis) after converting the negative and positive valence ratings to negative and positive numbers respectively. Error bars show adjusted 95% confidence intervals for repeated measures following the method proposed by Loftus and Masson (1994). The table below the bar chart describes valence means, standard deviations and 95% confidence intervals for each sound condition. The range for valence ratings was 1 (good aliens) to 5 (bad aliens). (b) Significant correlation between the total AQ score (x-axis; higher numbers indicate more symptoms of autism) and the composite rating for the consonant sound condition (y-axis; 1 = good aliens, 5 = bad aliens), wherein consonant sounds were judged as coming from more bad aliens as AQ scores increased. (c) (left) Items in the 'social skills' subscale of the Autism Spectrum Quotient (AQ) questionnaire (31). The 'social skills' items concentrate on mental state attribution processes (shown in italics) and avoidance of social interaction (all of the other items). (right) When all AQ subscales were included in the multiple regression equation, the social skills dimension was the best predictor for the valence rating of consonant sounds.

We examined the relationship between participants' valence attributions to the different sound conditions and self-reported symptoms of autism (Supplementary Table S8), which were assessed via the AQ questionnaire (Baron-Cohen et al. 2001).

A significant positive correlation was found between the AQ total score (AQTS) and the composite rating for consonant sounds (Pearson r_{37} = 0.434, P = 0.006; Figure 3b*)*, implying that participants with more autistic traits tended to rate consonant sounds as more negatively valenced. AQTS explained 19% of the variance in consonant sound rating (r^2 = 0.19). No significant correlations were found between the AQTS and the composite rating for strong dissonant sounds (Pearson r_{37} = 0.232, P = 0.115) or intermediate dissonant sounds (Pearson r_{37} = 0.169, P = 0.303). Importantly, these results showed that participants with more autistic traits were not evaluating all sounds more negatively, but only the consonant sounds.

A simple linear regression was further conducted to assess whether we could predict the level of accuracy in the rating for the consonant sounds from the AQTS. The level of accuracy was calculated by counting the number of correct responses ascribed to the consonant sounds (see methods). The results showed that the predictor variable AQTS contributed significantly to estimate the level of accuracy in participants' rating for the consonant sounds ($F_{1,37}$ = 7.228, P = 0.011; r^2 = 0.16; Figure 4), indicating that participants with higher levels of autistic traits presented difficulties in accurately recognising positive valence in consonant sounds.

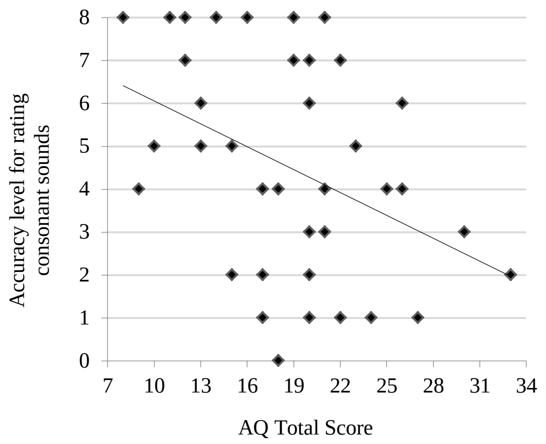


Fig. 4. Significant correlation between the AQ total score (AQTS) (x-axis; higher numbers indicate more symptoms of autism) and the level of accuracy in the rating for the consonant sounds (y-axis; 0 = none correct response, 8 = perfect accuracy), wherein the level of accuracy declined as AQ scores increased. Simple linear regression results showed that AQTS contributed significantly to estimate the level of accuracy in participants' rating for the consonant sounds. The derived equation for this relationship was: accuracy = 7.836 + -0.178 * AQTS. The ß-value of -0.178 represents the change in the outcome (accuracy) associated with a unit change in the predictor (AQTS). Therefore, if the predictor variable (AQTS) is, for instance, increased by 10 (representing more self reported symptoms of autism) the model predicts a decrease in accuracy of -1.78 (-0.178 * 10).

The AQ questionnaire contains five conceptually derived subscales (social skills, attention switching, attention to detail, communication, and imagination). A simultaneous multiple regression analysis was performed to investigate how well the combination of all subscales that integrate the AQTS predicted the rating for the consonant sounds.

[Checking assumptions for multiple regression analysis: For collinearity diagnostics we assessed the variance inflation factor (VIF; Myers, 1990) and tolerance statistics (Menard, 1995). In our model, the VIF values were well below 10, and the tolerance statistics were all well above 0.2; therefore, we can safely conclude that there was no collinearity within our data. We implemented a statistical measure to test for heteroscedasticity; employing the Breusch-Pagan test (Breusch and Pagan, 1979) we found that heteroscedasticity was not present in our model $F_{5,33}$ = 0.622, P = 0.684. (i.e. the assumption of homoscedasticity was met).]

The results of the multiple regression analysis were statistically significant $F_{5,33}$ = 2.88, P = 0.029 (n = 39). Both social skills and imagination correlated significantly with consonant sounds' rating. However, when all of the subscales were entered into the equation, only *social skills* emerged as a significant predictor (ß =

0.547, t_{38} = 2.70, P = 0.011; Spearman r_{37} = 0.626, P < 0.001); suggesting a strong link between autistic traits belonging to this dimension and biases in the emotional evaluation of consonant sound patterns (Figure 3c).

Relationship between valence ratings and a subset of AQ dimensions (social skills, communication, and imagination) as a proxy measure for social cognition.

The values for the multiple correlation coefficients between the social-cognition model (SCM predictors = social skills, communication, and imagination) and the outcome variable (valence ratings) showed a significant correlation between the SCM predictors and the composite rating for consonant sounds (Pearson r_{35} = 0.542, P = 0.006) (Table 6), reflecting that participants with more traits within this construct tended to rate consonant sounds as more negatively valenced. No significant correlations were found between the SCM predictors and the composite rating for strong dissonant sounds (Pearson r_{35} = 0.161, P = 0.817) or intermediate dissonant sounds (Pearson r_{35} = 0.275, P = 0.424).

The correlation matrix including consonant sounds' rating as outcome variable showed that there were no substantial correlations (r > 0.9) between predictors (i.e. no multicollinearity). Looking at the individual predictors, the highest correlation was between social skills and communication scales (r = 0.604, P < 0.001). Social skills and communication also correlated significantly with imagination (r = 0.455, P = 0.002 and r = 0.525, P < 0.001 respectively), suggesting that these predictors could be measuring a similar underlying factor. Of all the predictors, social skills correlated best with the rating for consonant sounds (r = 0.604, P < 0.001).

With the three social cognition predictors included, the amount of variability in the dependent variable (consonant sounds rating) accounted for by this model equalled 29.4% (r^2 = 0.294) (Table 6). Change statistics indicated that this r^2 change in the amount of variance explained was significant (F = 4.83, P = 0.006). The value for the Durbin-Watson statistic was 2.59, reflecting that the assumption of independent errors had almost certainly been met.

When performing a hierarchical regression, including a 2nd model in which attention switching and attention to detail were also incorporated (model 1 had the three social cognition predictors, model 2 had all five AQ predictors, consequently, the change in the number of predictors equals 2), the r^2 value only increased to 0.304 (30.4%), showing that the AQ dimensions of attention switching and attention to detail only accounted for an additional 1% of the variance, which gave rise to a non significant *F-ratio* of 0.244 (P = 0.785) (Table 6). These results suggest that only the initial (social cognition) model significantly improved our ability to predict the rating for the consonant sounds.

Table 6. Multiple correlation coefficients and change statistics for the specified models (Social cognition [model 1] predictors = social skills, communication, and imagination; Complete [model 2] predictors = [Model 1] + attention switching and attention to detail) and the dependent variable valence rating for consonant sounds.

Hierarchical regression	Multiple correlation coefficients		Change Statistics		
Models	r	r Square	r Square Change	*F* Change	*P* Change
1. Social cognition	0.542	0.294	0.294	4.853	0.006
2. Complete (incl. attention switching/detail)	0.551	0.304	0.010	0.244	0.785

Table 7. Model parameters: individual contribution of each of the predictors included in the social cognition model.

Social cognition model	Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for *b*	
Dimensions	*b-value*	Std. Error	ß	*t*	*P*	Lower Bound	Upper Bound
Social skills	0.139	0.051	0.5	2.749	0.009	0.036	0.242
Communication	-0.042	0.081	-0.1	-0.523	0.604	-0.206	0.121
Imagination	0.069	0.066	0.178	1.042	0.304	-0.065	0.204

Social cognition model parameters (Table 7): the b-values provide information on the relationship between the rating for consonant sounds and each social cognition predictor, in particular, they reflect to what degree each predictor affects the outcome variable, if the effects of all other predictors are held constant. Their associated standard errors indicate to what extent these values would vary across different samples, and are used to determine whether or not the b-values differ from zero. For the social cognition model, only the social skills dimension appears as a significant predictor of consonant sounds rating (t_{35} = 2.749, P < 0.001). Its *b*-value of 0.139 indicates that as traits in the social skills domain increase by one unit, the valence rating for consonant sounds increases by 0.139 units (i.e. more negative valence ascribed to consonant sounds). The standardized beta values (labeled as ß) inform the number of standard deviations that the rating for consonant sounds will change as a result of one standard deviation change in the predictor (since these values are measured in standard deviation units, they are directly comparable). For social skills (standardized ß = 0.50), which indicates that as traits in this dimension increase by one standard deviation (2.257), the valence rating for consonant sounds increases by 0.50 standard deviations. The standard deviation for the rating of consonant sounds is 0.628, and therefore this constitutes a change of 0.314 in the direction of negative valence (0.50 multiplied by 0.628). Finally, the confidence intervals of the unstandardized *b*-values are margins constructed such that in 95% of these samples these boundaries will contain the true value of *b*. For this model, social skills is the only predictor with confidence intervals that do not cross zero [95% CI for *b* = 0.036 – 0.242], its positive sign informs about the direction of the relationship between the predictor and the outcome variable, indicating that more negatively valenced judgements of consonants sounds would be expected with a higher number of traits in the social skills domain.

Discussion

The present study aimed to investigate the effects of sensory consonance/dissonance on the cognitive and neural mechanisms underlying the processing of valence during temporary inferences about others' intentions. The paradigm entailed a fine-grain emotion recognition task that required the intuitive ability to ascribe transitory intentional states in real time (also referred as 'online' mentalizing: Abell et al. 2000) based on non-verbal auditory cues. Participants had to categorize, in terms of positive/negative valence, auditory stimuli of distinct dissonance levels controlled by interval content.

Behavioural results consistent with previous evidence on valence judgments for consonant and dissonant sounds

Results revealed significant differences in valence inferences between consonant and strong dissonant sounds. Examination of the valence mean ratings for the three sound conditions showed that participants rated the sounds that consist of more consonant (dissonant) intervals as more positively (negatively) valenced, a result that was supported by the three experiments reported here. Participant's evaluation of consonances conveyed the fastest reaction times, and significantly faster than intermediate dissonances, which elicited the slowest reaction times. The mean valence rating for intermediate dissonant sounds yielded values between the strong dissonant and the consonant conditions in all experimental settings, but which could not be clearly discriminated from either of these contrasting conditions, rendering intermediate dissonances as the most ambiguous category. Taken together, behavioural results are consistent with previous research that has examined affective reactions induced by consonance/dissonance (Blood et al., 1999; Costa et al., 2000; Plomp and Levelt, 1965; Trainor and Heinmiller, 1998). We consider that the longer reaction times and uncertain valence response for the intermediate dissonances could relate to the characteristic ambiguity of this condition, which was constructed based on sequentially triggered minor thirds (building the content of a diminished seventh chord), a sonority which has been commonly applied to connote affective states of suspense and ambivalence in music (Huron, 2008; Meyer, 1956). With respect to tonalness, Temperley's Bayesian key-finding model (Temperley, 2010) situates this sound condition in-between the other two extreme conditions, further supporting its intermediate nature. This specific aspect may have probably affected participant's judgments at response selection rather than at perceptual processing stages.

Contextual memory systems supporting valence inferences

The neuroimaging experiment revealed increased activation of bilateral superior temporal regions when contrasting each sound condition against the baseline condition, with peak responses in primary auditory cortex (Heschl's gyri), reflecting basic response to auditory stimuli. Significant signal changes were also observed in the left parahippocampal cortex (lPHC). Several studies have indicated that the PHC may be involved in the emotional appraisal of varying levels of musical consonance/dissonance (Blood et al. 1999; Koelsch et al. 2006; Gosselin et al. 2006). We believe that this might be explained by its role in the storage and recall of long-term associations built up over repeated exposure (contextual memory: Aminoff et al. 2013). The culturally-acquired connotations ascribed to consonant and dissonant intervals may possibly be among these associations, together with other contextual elements that are necessary to define and bring meaning to the environment. We consider that associative memory could be a relevant source used to inform valence inferences during the task.

Modulatory effect of dissonance level on mental state attribution and attention reorienting processes

Distinct levels of consonance/dissonance exerted differential modulatory influences on the right temporo-parietal junction (rTPJ). Specifically, participants' inferences for signals consisting of highly dissonant intervals showed a stronger BOLD response in the rAG, a functionally-defined subregion of the rTPJ directly implicated in mental state attribution and inferences of transitory intentional states (Carter and Huettel, 2013; Van Overwalle, 2009).

We propose that the less coherent structure and negative valence associated with more dissonant auditory cues may have demanded heightened theory of mind processing resources. This hypothesis is consistent with the evidence provided by Young and collaborators, who found a stronger rTPJ response for negative judgments that rely on mental state information (Young et al. 2007; Young and Saxe, 2008; Young and Saxe, 2009a; Young and Saxe, 2009b, Young et al. 2010). These studies, however, were conducted employing verbal information (i.e. participants read stories that required inferences about a character's beliefs). Our findings therefore extend this evidence to the domain of spontaneous valence inferences based on non-verbal auditory stimuli acting as social affective cues.

Some researchers have argued that the 'theory of mind' theory of rTPJ functionality (Saxe and Kanwisher, 2003; Saxe and Wexler, 2005) may have neglected evidence showing rTPJ activity also in attentional paradigms that require reorienting (e.g. when volunteers must break their current attentional set to reorient to task relevant stimuli: Corbetta and Shulman, 2002), which are not specific to social contexts (Mitchell, 2008). Meta-analyses (Decety and Lamm, 2007) and direct comparisons between attention and theory of mind tasks (Mitchell, 2008) have revealed a generally overlapping brain response for both types of processing in the rTPJ. In particular, with regards to the present study, empirical evidence has also demonstrated an important role of the rAG during reorienting in exogenous cueing paradigms (Chambers et al. 2004). We believe that our findings could highlight a liaison between both theoretical accounts. Higher levels of dissonance may modulate attentional processes by leading participants to reorient their attention to specific non-verbal signs. From an evolutionary standpoint, there would clear advantages in directing attentional resources towards cues that could signal potentially threatening value in others' mental states (Corbetta et al. 2008; Carter and Huettel, 2013).

The engagement of the "salience network" (bilateral AI and ACC) (Menon and Uddin, 2010) for the comparison between strong and intermediate dissonances further supports the need to mark and segregate strong dissonant sounds for additional processing (possibly because of the intrinsic complexity and negative valence ascribed to dissonances) and appears to suggest, together with the response of the attention reorienting system, a demand for greater information integration when a stimulus is appraised as motivationally significant.

Relationship between participants' performance on the task and self-reported symptoms of autism

Several studies have demonstrated that individuals with autism process social affective cues differently. However, research directly assessing emotion recognition has, so far, reported contradictory results (Uljarevic and Hamilton, 2012). Most of these investigations have only examined emotions expressed by face and body, and, evidence suggests that these deficits might not generalize to the musical domain (Heaton et al. 1999; Allen et al. 2013). Some researchers, however, have proposed that more fine-grained and controlled tasks would be required to reveal the failure of these fundamental early emotion identification skills (Humphreys et al. 2007; Heaton et al. 2012). In Experiment 3, we investigated the relationship between participants' task performance and self-reported symptoms of autism in clinically typical adult participants with varying levels of autism spectrum traits (assessed via AQ questionnaire: Baron-Cohen et al. 2001). Notably, participants with higher levels of autistic traits presented significant difficulties and were less accurate in recognising positive valence in consonant sounds (Figure 3b and Figure 4). The statistical analysis demonstrated a strong association between autistic traits belonging to the social skills dimension and biases in the appraisal of consonant sounds (Figure 3c). We further assessed the relationship between valence ratings and a subset of AQ dimensions (social skills, communication, and imagination) as a proxy measure for social cognition. We found a significant correlation between the social cognition predictors and the composite rating for consonant sounds. Statistical analyses indicated that the social cognition predictors accounted for a significant amount of variance in the rating of consonant sounds, and further showed that the dimensions of attention switching and attention to detail did not account for a significant amount of variability in the valence judgments for this condition.

These findings do not represent an outcome that could have been easily anticipated by contemporary theories of autism. The "executive dysfunction" theory (Ozonoff et al., 1994; Russell, 1998) would have predicted an overall impaired performance, since the response to all sound categories requires executive function (i.e. decision making and handling a novel and technically difficult task outside the domain of automatic psychological processes). The results, however, did not yield a global impairment but one specific to the judgement of the consonant condition. The "empathizing-systemizing" theory (E-S) (Baron-Cohen et al.,

2005) explains systemizing as the drive to analyse the variables in a system to derive the underlying rules that govern its behaviour (Baron-Cohen et al., 2003). The systemizing quotient includes items such as: ‘in maths, I am intrigued by the rules and patterns governing numbers’ / ‘when I listen to a piece of music, I always notice the way it’s structured’, etc (Baron-Cohen et al., 2002). The E-S theory of autism predicts impaired empathizing and normal (i.e. in line with mental age) or even superior systemizing if a domain is systemizable. Since music is a systemizable domain according to this theory (Baron-Cohen et al., 2002), a general accurate rating to all sound conditions would have been expected due to the normal or superior systemizing capabilities evidenced in individuals with autism symptoms (Baron-Cohen et al., 2002). Such prediction was not substantiated in this study. The E-S theory does predict that non-systemizables domains like fiction would be poorly integrated in autism. Accordingly, it could have been speculated that the ‘fictional’ task demanded in this experiment might itself impact on participants’ conserved systemizing capacities. However, the theory would still not explain why the results did not reveal a general inappropriate valence attribution, but only circumscribed to the consonant category. Lastly, both the “weak central coherence” theory (Frith, 1989; Happé, 1996), with its emphasis on local bias and global impairment, and the “enhanced perceptual functioning” theory (Mottron et al., 2006), which highlight superior low-level perception abilities (e.g. superior pitch processing ability), would have predicted an accurate task performance for all sound conditions based on the proposed detail focus and enhanced low-level discrimination characteristic of the ASD perceptual-cognitive style, which, as described, was neither evidenced.

The results appear to be in agreement with previous research, which has shown that although individuals with autistic symptoms may be able to categorize certain types of stimuli during mentalizing tasks (i.e. correctly identify whether mental states are involved, and even use mental state language to describe these stimuli), they can manifest significant difficulties in accurately understanding the feelings or emotions represented by specific patterns of stimuli, to which they may attribute inappropriate mental states (Abell et al. 2000; Castelli et al. 2002). Humphreys et al. (2007) also observed deficits in the recognition of positive emotions in a face emotion recognition task through employing stimuli presented at moderate intensities (the authors found that the individuals with autism were significantly less likely to identify blends containing 50%

happiness as happy, compared to the control group). Our findings, however, do not point towards a simple negative processing bias. According to cognitive models, not only positive but also ambiguous events are subject to this bias (Krantz and Rude, 1984; Hollon et al. 1986). On the contrary, our study revealed a specific pattern of misinterpreting the stimuli, which only affected the appraisal of positive cues (consonant condition), but which did not affect the evaluation of ambiguous stimuli (i.e. intermediate dissonance condition). The difficulty evidenced appears to be linked to social cognition processes dealing with cues that would potentially signal (and therefore enable) a positive social interaction.

The analysis of the neural mechanisms engaged while participants were judging consonant sounds revealed a marked coactivation between the lPHC and the rAG. Given the involvement of the PHC in the storage and recall of long-term associations (the memory that binds different items together; e.g. consonant sounds cue retrieval of positive valence associations: Gosselin et al. 2006; Aminoff et al. 2013), and the role of the rAG in reasoning about others' mental states and, specifically, in the representation of temporary intentions of an actor (Keysers and Perrett, 2004; Saxe and Wexler, 2005; Keysers and Gazzola, 2007; Van Overwalle, 2009; Van Overwalle and Baetens, 2009), we interpreted these findings to imply an active process in which contextual associations for the consonant sound patterns (e.g. pleasantness, smoothness) were being retrieved to emotionally qualify/inform the inferences represented at the level of the rAG (*good-friendly* aliens). Effective connectivity analysis further demonstrated a directional transfer of information in which inhibitory feedback signals were being sent from the lPHC to the rAG, substantiating our prediction about deactivations in theory of mind substrates modulated by consonances. Dissonances, on the other hand, demanded not only enhanced mental state/attention reorienting processing resources but also greater information integration. In light of these results, and following the highly specific difficulties in the recognition of positive valence in consonant sounds observed for participants with more autistic traits, we think that a compromised use of associative information (based on previous experience) to contextualize temporary intention inferences might underlie these findings. A convergent neurobiological formulation has been recently proposed by Lawson, Rees and Friston (2014) following the view articulated by Pellicano and Blur (2012), they argued that abnormalities in autistic perception and social interaction could arise from a failure to contextualize sensory cues in relation to prior beliefs stored in memory systems (i.e. autistic observers seem to be less influenced by contextual information). We therefore consider that the investigation of the functional link between associative memory and mental state attribution/attention reorienting (rTPJ) might provide insights that could improve the understanding of some aspects of autistic symptomatology (social and communication factors). Following our findings, we believe that a relevant research direction concerns the analysis of afferent modulatory influences on temporary mental state inferences and, specifically, the assessment of whether the interaction between the network of areas that processes contextual associations (i.e. medial prefrontal, retrospenial and parahippocampal cortices) and the rTPJ, differ between individuals with ASD and matched controls. Altered functioning of integrative circuitry at the level of the rTPJ has been recently reported in individuals with high functioning autism (for alterations in fractional anisotropy properties see Thakkar et al. 2008; for abnormal white matter volume see McAlonan et al. 2009; for decreased functional connectivity see Castelli et al. 2002; Cherkassky et al. 2006; Just et al. 2007; Mueller et al. 2013). However, this study was conducted with typically developing individuals (i.e. none of the participants tested here had a diagnosis of autism, subjects were clinically typical individuals who varied in their levels of autism traits measured by the AQ questionnaire). Research is envisioned that examines the hypothesis hereby presented in prospective behavioural and neuroimaging studies conducted with individuals clinically diagnosed with autism.

Considerations for future research with individuals diagnosed with autism

This subsection describes certain aspects that should be taken into consideration for prospective research. The term “autism spectrum disorder” (ASD), as currently defined by the American Psychiatric Association (APA, 2013; DSM-5), covers a wide range of neurodevelopmental disorders including: ‘Autistic disorder’, ‘Asperger’s disorder’, ‘Childhood disintegrative disorder’ and ‘Pervasive developmental disorder not otherwise specified’. The Neurodevelopmental Work Group from APA has considered that “autism spectrum disorder”, as a single umbrella disorder “will improve the diagnosis of ASD without limiting the sensitivity of the criteria”. The symptoms of individuals with ASD fall on a continuum, with some individuals showing mild symptoms and others having much more severe symptoms. The disorder is currently diagnosed on the basis of social communication and interaction deficits (e.g. responding inappropriately in conversations, misreading nonverbal interactions, or having difficulty building friendships appropriate to their age), and the presence of restricted and repetitive patterns of behaviour, interests or activities.

It is important to note that the impairment in autism has been specifically linked to the social communication and interaction domain (APA, 2013; DSM-5), and not to emotion processing per se. In accordance with the deficits in social communication and interaction, described as clinical indicators of autism, impaired mentalizing or theory of mind has been observed in numerous studies [for a review see (Baron-Cohen, 2000)]. With reference to the difficulties in the emotional domain, which have been historically considered a feature of autism (Asperger, 1944; Kanner, 1968), empirical studies examining emotion recognition have reported contradictory findings [for reviews see: (Begeer et al., 2008; Harms et al., 2010; Uljarevic and Hamilton, 2013)]. For example, Williams and Happé (2010) found that children with autism were as able as age and ability-matched comparison participants to recognise ‘social’ and ‘non-social’ emotions in others, and to describe their own previous experiences of these emotions. Although similar performance levels between the two groups were expected with regard to the recognition and reporting of non-social emotions, such similar levels of understanding of social emotions were not predicted considering the counter-evidence reported in previous studies (Capps et al., 1992; Heerey et al., 2003; Losh and Capps, 2006). This was not the only study that failed to find impaired recognition of social emotions in autism (Allen et al., 2013; Castelli, 2005; Grossman et al., 2000; Heaton et al., 1999, 2008; Jones et al., 2011). Moreover, a close reading of the empirical literature reveals that the claim of global emotional difficulties in autism can be markedly inconsistent (Begeer et al., 2008; Harms et al., 2010; Uljarevic and Hamilton, 2013). As Geoff Bird has stated, “more consistent is the substantial variability within the population of individuals with autism” (Bird and Cook, 2013).

Pamela Heaton and collaborators have provided valuable information about three factors that may strongly influence the performance on emotion recognition tasks in ASD (Heaton et al., 2012). A first factor is stimulus complexity. Studies employing stimulus morphing to achieve subtle varying levels in emotional intensity have revealed that deficits in the ASD group emerged only when higher levels of neutral were incorporated (i.e. not ‘full blown’ or ‘100% expression’ stimuli) and the emotions portrayed became less intense (Humphreys et al., 2007; Montagne et al., 2007). A second factor that could influence the development and affect emotion recognition skills is sensory dysfunction. Studies have suggested that individuals with autism might also have deficits in basic perceptual abilities including multisensory processing [behavioural evidence reported in (Foss-Feig et al., 2010), neural evidence reported in (Russo et al., 2008), for a review of neurophysiologic findings see (Marco et al., 2011)]. In this sense, abnormalities in sensory processing may impact on the autistic infant's ability to learn about facial and vocal expressions of emotion (Kern et al., 2006). A third factor concerns alexithymia. Bird and collaborators have proposed what they termed "the alexithymia hypothesis" to suggest that, where observed, the emotional symptoms of autism could be specifically due to the greater proportion of individuals with alexithymia in the autistic population (Bird and Cook, 2013). Alexithymia is a disorder characterised by reduced or absent affective responses (type I alexithymia) or difficulties in understanding and ascribing affective labels to one's own bodily sensations of emotional arousal, even when affective arousal is present (type II alexithymia) (Bermond et al., 2007). Studies have demonstrated that levels of alexithymia are greatly increased in the ASD populations (Berthoz and Hill, 2005; Bird et al., 2010, 2011; Hill et al., 2004; Silani et al., 2008). Importantly, recent empirical evidence has demonstrated associations between high levels of alexithymia and increased difficulties in emotion recognition [for associations between alexithymia and difficulties in understanding expressions of emotions in faces and voices see (Heaton et al., 2012); for associations between alexithymia and emotional responses to music see (Allen et al., 2013)]. Taken together, these findings emphasise the importance of considering stimulus complexity when designing emotion materials, and highlight the critical value of inspecting sensory processing skills and evaluating levels of alexithymia in prospective studies of emotion recognition conducted with individuals clinically diagnosed with ASD.

Limitations of the present study

Although past neuroscientific experiments have already studied how humans draw on "social intelligence" to ascribe temporary intentions or goals to an actor (enabling social inferences such as: Is the other person friendly or aggressive?) and concluded that the rTPJ encodes representations about these intuitive inferences (Keysers and Perrett, 2004; Saxe and Wexler, 2005; Keysers and Gazzola, 2007; Van Overwalle, 2009; Van

Overwalle and Baetens, 2009); a limitation of the present experiment was the lack of a supplementary control condition that could have enabled a precise identification of 'theory of mind' neural indices specific for our task. The reason for this is that our study did not aim to provide neuroimaging evidence for the specificity of participants' judgments within the domain of mental attribution, but to investigate the impact of sensory consonance/dissonance level on the neural substrates underlying *valence* inference processes. Participants' task performance was deemed to rely on intentional processing considering previous literature related to the specific question that subjects were required to respond, which involved the attribution of behavioral intentions to others based on non-verbal cues (Brunet et al. 2000; den Ouden et al., 2005; Saxe and Wexler, 2005; Blackemore et al., 2007).

Conclusions

Overall, our findings show that the controlled and systematic manipulation of musical structural features within social cognition paradigms can be applied to deepen our understanding of the functional anatomy underlying music-evoked emotions. Importantly, a precise delineation of the brain mechanisms involved might provide a model system for characterising neural networks during emotion processing, which could prove valuable for the investigation of abnormal functional integration in disease states.

Acknowledgements

We thank S. Koelsch and M. Rohrmeier for helpful discussions. We thank the staff of Fundación Científica del Sur Imaging Centre (technicians: D. Sarroca, V. Maciel and G. Serruto) for help with data acquisition and C. Morales for his assistance with neuroanatomical analyses. This project was made possible through the support of a Queens' College Cambridge Walker Studentship to F. Bravo. Financial support has also been provided by: Zukunftskonzept at TU Dresden (Exzellenzinitiative of the Deutsche Forschungsgemeinschaft), SEMPRE (Arnold Bentley New Initiatives Fund), and Cambridge Society for the Application of Research through a CSAR Student Award to F. Bravo.

References

Abell, F., Happé, F., & Frith, U. (2000) Do triangles play tricks? Attributions of mental states to animated shapes in normal and abnormal development. Journal of Cognition and Development, 15, 1-16.

Adolphs, R., Sears, L., & Piven, J. (2001) Abnormal processing of social information from faces in autism. Journal of Cognitive Neuroscience, 13, 232-240.

Alink, A., Schwiedrzik, C. M., Kohler, A., Singer, W., & Muckli, L. (2010). Stimulus Predictability Reduces Responses in Primary Visual Cortex. Journal of Neuroscience, 30(8), 2960–2966. https://doi.org/10.1523/JNEUROSCI.3730-10.2010
Allen, R., Davis, R., & Hill, E. (2013) The effects of autism and alexithymia on physiological and verbal responsiveness to music. Journal of Autism and Developmental Disorders, 43, 432-444.
American Psychiatric Association. (2013) Diagnostic and statistical manual of mental disorders: DSM-5. Washington, DC: Author.
Aminoff, E. M., Kveraga, K., & Bar, M. (2013) The role of the parahippocampal cortex in cognition. Trends in Cognitive Science, 17(8), 379-390.
Apel, W. (1972) Harvard Dictionary of Music. Cambridge: Harvard Univ Press.
Asperger, H. (1944). Die „Autistischen Psychopathen" im Kindesalter. Archiv für Psychiatrie und Nervenkrankheiten, 117(1), 76–136. https://doi.org/10.1007/BF01837709
Astington, J., Harris, P., & Olson, D. (1988) Developing Theories of Mind. Cambridge, UK: Cambridge Univ Press.
Ayres, T., Aeschbach, S., & Walker, E. L. (1980). Psychoacoustic and experiential determinants of tonal consonance. The Journal of Auditory Research, 20(1), 31–42.
Bar, M. (2004). Visual objects in context. Nature Reviews Neuroscience, 5(8), 617–629. https://doi.org/10.1038/nrn1476
Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. Trends in Cognitive Sciences, 11(7), 280–289. https://doi.org/10.1016/j.tics.2007.05.005
Bar, M., & Ullman, S. (1996). Spatial context in recognition. Perception, 25(3), 343–352. https://doi.org/10.1068/p250343
Bar, M., Aminoff, E., & Schacter, D. L. (2008). Scenes unseen: The parahippocampal cortex intrinsically subserves contextual associations, not scenes or places per se. The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 28(34), 8539–8544. https://doi.org/10.1523/JNEUROSCI.0987-08.2008
Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., et al. (2006) Top-down facilitation of visual recognition. Proceedings of the National Academy of Sciences, U.S.A., 103, 449-454.
Baron-Cohen, S. (2000). Theory of mind and autism: A review. In International Review of Research in Mental Retardation (Vol. 23, pp. 169–184). Elsevier. https://doi.org/10.1016/S0074-7750(00)80010-5
Baron-Cohen, S., & Cross, P. (1992) Reading the eyes: Evidence for the role of perception in the development of a theory of mind. Mind & Language, 6, 173-186.
Baron-Cohen, S., Richler, J., Bisarya, D., Gurunathan, N., & Wheelwright, S. (2003). The systemizing quotient: an investigation of adults with Asperger syndrome or high-functioning autism, and normal sex differences. Philosophical Transactions of the Royal Society B: Biological Sciences, 358(1430), 361–374. https://doi.org/10.1098/rstb.2002.1206
Baron-Cohen, S., Wheelwright, S., Griffin, R., Lawson, J. & Hill, J. (2002). The exact mind: empathising and systemising in autism spectrum conditions. In Handbook of cognitive development (ed. U. Goswami). Oxford: Blackwell.
Baron-Cohen, S., Wheelwright, S., Lawson, J., Griffin, R., Ashwin, C., Billington, J., & Chakrabarti, B. (2005). Empathizing and systemizing in autism spectrum conditions. Handbook of Autism and Pervasive Developmental Disorders, 1, 628–639.
Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001) The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. Journal of Autism and Developmental Disorders, 31, 5-17.
Barrett, L. F., & Wager, T. D. (2006). The structure of emotion: Evidence from neuroimaging studies. Current Directions in Psychological Science, 15(2), 79–83.
Begeer, S., Koot, H. M., Rieffe, C., Meerum Terwogt, M., & Stegge, H. (2008). Emotional competence in children with autism: Diagnostic criteria and empirical evidence. Developmental Review, 28(3), 342–369. https://doi.org/10.1016/j.dr.2007.09.001
Bermond, B., Clayton, K., Liberova, A., Luminet, O., Maruszewski, T., Bitti, P. E. R., … Wicherts, J. (2007). A cognitive and an affective dimension of alexithymia in six languages and seven populations. Cognition and Emotion, 21(5), 1125–1136. https://doi.org/10.1080/02699930601056989
Berthoz, S., & Hill, E. L. (2005). The validity of using self-reports to assess emotion regulation abilities in adults with autism spectrum disorder. European Psychiatry, 20(3), 291–298. https://doi.org/10.1016/j.eurpsy.2004.06.013
Bharucha, J. J., & Pryor, J. H. (1986) Disrupting the isochrony underlying rhythm: an asymmetry in discrimination. Perception & Psychophysics Journal, 40, 137-141.
Bidelman, G. M., & Krishnan, A. (2011). Brainstem correlates of behavioral and compositional preferences of musical harmony. Neuroreport, 22(5), 212–216. https://doi.org/10.1097/WNR.0b013e328344a689
Bigand, E., & Parncutt, R. (1999) Perception of musical tension in long chord sequences. Psychological Research, 62, 237-254.
Bigand, E., Parncutt, R., & Lerdahl, F. (1996) Perception of musical tension in short chord sequences: The influence of harmonic function, sensory dissonance, horizontal motion, and musical training. Perception & Psychophysics Journal, 58, 124-141.
Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where Is the Semantic System? A Critical Review and Meta-Analysis of 120 Functional Neuroimaging Studies. Cerebral Cortex, 19(12), 2767–2796. https://doi.org/10.1093/cercor/bhp055

Bird, G., & Cook, R. (2013). Mixed emotions: the contribution of alexithymia to the emotional symptoms of autism. Translational Psychiatry, 3(7), e285. https://doi.org/10.1038/tp.2013.61
Bird, G., Press, C., & Richardson, D. C. (2011). The Role of Alexithymia in Reduced Eye-Fixation in Autism Spectrum Conditions. Journal of Autism and Developmental Disorders, 41(11), 1556–1564. https://doi.org/10.1007/s10803-011-1183-3
Bird, G., Silani, G., Brindley, R., White, S., Frith, U., & Singer, T. (2010). Empathic brain responses in insula are modulated by levels of alexithymia but not autism. Brain, 133(5), 1515–1525. https://doi.org/10.1093/brain/awq060
Blakemore, J.S., den Ouden, H., Choudhury, S., Frith, C. (2007) Adolescent development of the neural circuitry for thinking about intentions. Social Cognitive Affective Neuroscience, 2, 130-9.
Blood, A. J., Zatorre, R. J., Bermudez, P., & Evans, A. C. (1999) Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions. Nature Neuroscience, 2, 382-387.
Boltz, M. G. (2001) Musical soundtracks as a schematic influence on the cognitive processing of filmed events. Music Perception, 18, 427-454.
Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. Journal of Behavior Therapy and Experimental Psychiatry, 25(1), 49-59.
Bravo, F. (2012) The influence of music on the emotional interpretation of visual contexts. In M. Aramaki, M. Barthet, R. Kronland-Martinet, & S. Ystad (Eds.), From sounds to music and emotions (pp. 366-377). Berlin: Springer.
Bravo, F. (2014) Changing the interval content of algorithmically generated music changes the emotional interpretation of visual images. In M. Aramaki, M. Derrien, R. Kronland-Martinet, & S. Ystad (Eds), Sound, music and motion (pp. 494-508). Berlin: Springer.
Brunet, E., Sarfati, Y., Hardy-Baylé, M.C., & Decety, J. (2000) A PET Investigation of the Attribution of Intentions with a Nonverbal Task. NeuroImage, 11(2), 157-166.
Bugg, E. G. (1970). An Experimental Study of Factors Influencing Consonance Judgments. Johnson.
Burns, E. M. (1999). 7 - Intervals, Scales, and Tuning*. In D. Deutsch (Ed.), The Psychology of Music (Second Edition) (pp. 215–264). San Diego: Academic Press. https://doi.org/10.1016/B978-012213564-4/50008-1
Capps, L., Yirmiya, N., & Sigman, M. (1992). Understanding of Simple and Complex Emotions in Non-retarded Children with Autism. Journal of Child Psychology and Psychiatry, 33(7), 1169–1182. https://doi.org/10.1111/j.1469-7610.1992.tb00936.x
Carrasco, M. (2014). Spatial covert attention: Perceptual modulation. The Oxford Handbook of Attention, 183–230.
Carrasco, M., Williams, P. E., & Yeshurun, Y. (2002). Covert attention increases spatial resolution with or without masks: support for signal enhancement. Journal of Vision, 2(6), 467–479. https://doi.org/10.1167/2.6.4
Carter, R.M., & Huettel, S. A. (2013) A nexus model of the temporal–parietal junction. Trends in Cognitive Sciences, 17(7), 328-336.
Castelli, F. (2005). Understanding emotions from standardized facial expressions in autism and normal development. Autism, 9(4), 428–449. https://doi.org/10.1177/1362361305056082
Castelli, F., Frith, C., Happé, F., & Frith, U. (2002) Autism Asperger Syndrome and brain mechanisms for the attribution of mental states to animated shapes. Brain, 125, 1839-1849.
Chambers, C. D., Payne, J. M., Stokes, M. G., & Mattingley, J. B. (2004) Fast and slow parietal pathways mediate spatial attention. Nature Neuroscience, 7, 217-218.
Cherkassky, V. L., Kana, R. K., Keller, T. A., & Just, M. A. (2006) Functional connectivity in a baseline resting-state network in autism. Neuroreport, 17, 1687-1690.
Chiandetti, C., & Vallortigara, G. (2011). Chicks like consonant music. Psychological Science, 22(10), 1270–1273. https://doi.org/10.1177/0956797611418244
Cohen, A. J. (2001) Music as a source of emotion in film. In P. Juslin & J. Sloboda (Eds.), Handbook of Music and Emotion: Theory, Research, Applications (pp. 249-272). Oxford: Oxford University Press.
Colombetti, G. (2005). Appraising valence. Journal of Consciousness Studies, 12(8–9), 103–126.
Corbetta, M., & Shulman, G. L. (2002) Control of goal-directed and stimulus-driven attention in the brain. Nature Reviews Neuroscience, 3, 201-215.
Corbetta, M., Patel, G., & Shulman, G. L. (2008) The reorienting system of the human brain: from environment to theory of mind. Neuron, 58, 306-324.
Corbetta, M., Patel, G., and Shulman, G. L. (2008) The reorienting system of the human brain: from environ- ment to theory of mind. Neuron 58, 306-324.
Costa, M., Bitti, P. E. R., & Bonfiglioli, L. (2000). Psychological connotations of harmonic musical intervals. Psychology of Music, 28(1), 4–22.
Cousineau, M., McDermott, J. H., & Peretz, I. (2012). The basis of musical consonance as revealed by congenital amusia. Proceedings of the National Academy of Sciences, 109(48), 19858–19863. https://doi.org/10.1073/pnas.1207989109

Dalton, K. M., Nacewicz, B. M., Johnstone, T., Schaefer, H. S., Gernsbacher, M. A., Goldsmith, H. H., et al. (2005) Gaze fixation and the neural circuitry of face processing in autism. Nature Neuroscience, 8(4), 519-526.
Decety, J., & Lamm, C. (2007) The role of the right temporoparietal junction in social interaction: how low-level computational processes contribute to meta-cognition. Neuroscientist, 13, 580-593.
den Ouden, H.E.M, Frith, U., Frith, C., Blakemore, S.J. (2005) Thinking about intentions. NeuroImage, 28, 787–96.
DeWitt, L. A., & Crowder, R.G. (1987) Tonal fusion of consonant musical intervals. Perception & Psychophysics, 41, 73-84.
Dosher, B. A., & Lu, Z.-L. (2000). Noise exclusion in spatial attention. Psychological Science, 11(2), 139–146.
Dowling, W. J., & Harwood, D.L. (1986) Music cognition. San Diego: Academic Press.
Dumoulin, S. O., & Hess, R.F. (2007) Cortical specialization for concentric shape processing. Vision Research, 47, 1608-1613.
Ekman, P., & Friesen, W. V. (1971) Constants across cultures in the face and emotion. Journal of Personality and Social Psychology, 17, 124-129.
Enns, J. T., & Lleras, A. (2008). What's next? New evidence for prediction in human vision. Trends in Cognitive Sciences, 12(9), 327–333. https://doi.org/10.1016/j.tics.2008.06.001
Escoffier, N., Zhong, J., Schirmer, A., & Qiu, A. (2013). Emotional expressions in voice and music: Same code, same effect? Human Brain Mapping, 34(8), 1796–1810. https://doi.org/10.1002/hbm.22029
Fang, F., Kersten, D., & Murray, S.O. (2008) Perceptual grouping and inverse fMRI activity patterns in human visual cortex. Journal of Vision, 8, 1-9.
Fannin, H. A., & Braud, W. G. (1971). Preference for Consonant over Dissonant Tones in the Albino Rat. Perceptual and Motor Skills, 32(1), 191–193. https://doi.org/10.2466/pms.1971.32.1.191
Fletcher, P. C., Happé, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S. J., & Frith, C. D. (1995). Other minds in the brain: a functional imaging study of "theory of mind" in story comprehension. Cognition, 57(2), 109–128. https://doi.org/10.1016/0010-0277(95)00692-R
Foss-Feig, J. H., Kwakye, L. D., Cascio, C. J., Burnette, C. P., Kadivar, H., Stone, W. L., & Wallace, M. T. (2010). An extended multisensory temporal binding window in autism spectrum disorders. Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale, 203(2), 381–389. https://doi.org/10.1007/s00221-010-2240-4
Foss, A. H., Altschuler, E. L., & James, K. H. (2007). Neural correlates of the Pythagorean ratio rules. Neuroreport, 18(15), 1521–1525.
Frances, R. (1972) La perception de la musique. Paris, France: Librairie Philosophique 1.
Frijda, N. H. (1986). The Emotions. Cambridge University Press.
Friston, K. (2005). A theory of cortical responses. Philosophical Transactions of the Royal Society B: Biological Sciences, 360(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622
Friston, K. J. (2009) The free-energy principle: a rough guide to the brain? Trends in Cognitive Sciences, 13(7), 293-301.
Friston, K. J., Ashburner, J., Frith, C. D., Poline, J-B., Heather, J. D., & Frackowiak, R. S. J. (1995) Spatial registration and normalization of images. Human Brain Mapping, 2, 165-189.
Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997) Psychophysiological and modulatory interactions in neuroimaging. Neuroimage, 6, 218-229.
Friston, K. J., Harrison, L., & Penny, W. (2003) Dynamic causal modelling. Neuroimage, 19, 1273-1302.
Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. Philosophical Transactions of the Royal Society B: Biological Sciences, 364(1521), 1211–1221. https://doi.org/10.1098/rstb.2008.0300
Frith, U. (1989). Autism: explaining the enigma. Oxford, UK; Cambridge, MA, USA: Basil Blackwell.
Frith, U., & Frith, C. D. (2003) Development and neurophysiology of mentalizing. Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 358, 459-473.
Fritz, T. H., Renders, W., Müller, K., Schmude, P., Leman, M., Turner, R., & Villringer, A. (2013). Anatomical differences in the human inferior colliculus relate to the perceived valence of musical consonance and dissonance. European Journal of Neuroscience, n/a-n/a. https://doi.org/10.1111/ejn.12305
Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., … Koelsch, S. (2009). Universal Recognition of Three Basic Emotions in Music. Current Biology, 19(7), 573–576. https://doi.org/10.1016/j.cub.2009.02.058
Fujisawa, T. X., & Cook, N. D. (2011). The perception of harmonic triads: an fMRI study. Brain Imaging and Behavior, 5(2), 109–125. https://doi.org/10.1007/s11682-011-9116-5
Gallagher, H. L., Happé, F., Brunswick, N., Fletcher, P. C., Frith, U., & Frith, C. D. (2000). Reading the mind in cartoons and stories: an fMRI study of "theory of mind"in verbal and nonverbal tasks. Neuropsychologia, 38(1), 11–21.
Gobbini, M. I., Koralek, A. C., Bryan, R. E., Montgomery, K. J., & Haxby, J. V. (2007). Two Takes on the Social Brain: A Comparison of Theory of Mind Tasks. Journal of Cognitive Neuroscience, 19(11), 1803–1814. https://doi.org/10.1162/jocn.2007.19.11.1803

Gosselin, N., Samson, S., Adolphs, R., Noulhiane, M., Roy, M., Hasboun, D., et al. (2006) Emotional responses to unpleasant music correlates with damage to the parahippocampal cortex. Brain, 129, 2585-2592.
Green, A. C., Baerentsen, K. B., Stødkilde-Jørgensen, H., Wallentin, M., Roepstorff, A., & Vuust, P. (2008). Music in minor activates limbic structures: a relationship with dissonance? Neuroreport, 19(7), 711–715. https://doi.org/10.1097/WNR.0b013e3282fd0dd8
Grossberg, S. (2009). Cortical and subcortical predictive dynamics and learning during perception, cognition, emotion and action. Philosophical Transactions of the Royal Society B: Biological Sciences, 364(1521), 1223–1234. https://doi.org/10.1098/rstb.2008.0307
Grossman, J. B., Klin, A., Carter, A. S., & Volkmar, F. R. (2000). Verbal Bias in Recognition of Facial Emotions in Children with Asperger Syndrome. Journal of Child Psychology and Psychiatry, 41(3), 369–379. https://doi.org/10.1111/1469-7610.00621
Gusnard, D., & Raichle, M. E. (2001) Searching for a baseline: functional imaging and the resting human brain. Nature Reviews Neuroscience, 2, 685-694.
Happé, F. G. E. (1996). Studying Weak Central Coherence at Low Levels: Children with Autism do not Succumb to Visual Illusions. A Research Note. Journal of Child Psychology and Psychiatry, 37(7), 873–877. https://doi.org/10.1111/j.1469-7610.1996.tb01483.x
Harms, M. B., Martin, A., & Wallace, G. L. (2010) Facial emotion recognition in autism spectrum disorders: a review of behavioral and neuroimaging studies. Neuropsychology Review, 20(3), 290-322.
Heaton, P., Allen, R., Williams, K., Cummins, O., & Happé, F. (2008). Do social and cognitive deficits curtail musical understanding? Evidence from autism and Down syndrome. British Journal of Developmental Psychology, 26(2), 171–182. https://doi.org/10.1348/026151007X206776
Heaton, P., Hermelin, B., & Pring, L. (1999) Can children with autistic spectrum disorders perceive affect in music? An experimental investigation. Psychological Medicine, 29, 1405-1410.
Heaton, P., Reichenbacher, L., Sauter, D., Allen, R., Scott, S., & Hill, E. (2012) Measuring the effects of alexithymia on perception of emotional vocalizations in autistic spectrum disorder and typical development. Psychological Medicine, 42, 2453-2459.
Heerey, E. A., Keltner, D., & Capps, L. M. (2003). Making Sense of Self-Conscious Emotion: Linking Theory of Mind and Emotion in Children With Autism. Emotion, 3(4), 394–400. https://doi.org/10.1037/1528-3542.3.4.394
Heider, F., & Simmel, M. (1944). An Experimental Study of Apparent Behavior. The American Journal of Psychology, 57(2), 243. https://doi.org/10.2307/1416950
Helmholtz, H. von, & Ellis, A. J. (1895). On the sensations of tone as a physiological basis for the theory of music. London, New York : Longmans, Green, and Co. (Original German work published 1863).
Hill, E., Berthoz, S., & Frith, U. (2004). Brief report: cognitive processing of own emotions in individuals with autistic spectrum disorder and in their relatives. Journal of Autism and Developmental Disorders, 34(2), 229–235.
Hollon, S. D., Kendall, P. C., & Lumry, A. (1986) Specificity of depressotypic cognitions in clinical depression. Journal of Abnormal Psychology, 95, 52-59.
Humphreys, K., Minshew, N., Leonard, G. L., & Behrmann, M. (2007) A fine-grained analysis of facial expression processing in high-functioning adults with autism. Neuropsychologia, 45, 685-695.
Huron, D. (2008). Sweet Anticipation: Music and the Psychology of Expectation. Cambridge, Mass.: A Bradford Book.
Itoh, K., Suwazono, S., & Nakada, T. (2010). Central auditory processing of noncontextual consonance in music: an evoked potential study. The Journal of the Acoustical Society of America, 128(6), 3781–3787. https://doi.org/10.1121/1.3500685
Izumi, A. (2000). Japanese monkeys perceive sensory consonance of chords. The Journal of the Acoustical Society of America, 108(6), 3073–3078.
Jones, C. R. G., Pickles, A., Falcaro, M., Marsden, A. J. S., Happé, F., Scott, S. K., … Charman, T. (2011). A multimodal approach to emotion recognition ability in autism spectrum disorders. Journal of Child Psychology and Psychiatry, and Allied Disciplines, 52(3), 275–285. https://doi.org/10.1111/j.1469-7610.2010.02328.x
Just, M. A., Cherkassky, V. L., Keller, T. A., Kana, R. K., & Minshew, N. J. (2007) Functional and anatomical cortical underconnectivity in autism: evidence from an fMRI study of an executive function task and corpus callosum morphometry. Cerebral Cortex, 17, 951-961.
Kahan, J., & Foltynie, T. (2013). Understanding DCM: Ten simple rules for the clinician. NeuroImage, 83, 542–549. https://doi.org/10.1016/j.neuroimage.2013.07.008
Kameoka, A., & Kuriyagawa, M. (1969) Consonance theory. II. Consonance of complex tones and its calculation method. The Journal of the Acoustical Society of America, 45, 1460-1469.
Kanner, L. (1968). Autistic disturbances of affective contact. Acta Paedopsychiatrica, 35(4), 100–136.
Kern, J. K., Trivedi, M. H., Garver, C. R., Grannemann, B. D., Andrews, A. A., Savla, J. S., … Schroeder, J. L. (2006). The pattern of sensory processing abnormalities in autism. Autism, 10(5), 480–494. https://doi.org/10.1177/1362361306066564

Keysers, C., & Gazzola, V. (2007) Integrating simulation and theory of mind: From self to social cognition. Trends in Cognitive Sciences, 11, 194-6.

Keysers, C., & Perrett, D. I. (2004) Demystifying social cognition: A Hebbian perspective. Trends in Cognitive Sciences, 8, 501-7.

Koelsch, S. (2011) Towards a neural basis of processing musical semantics. Physics of Life Reviews, 8, 89-105.

Koelsch, S., Fritz, T., Cramon, D. Y., Muller, K., & Friedereci, A. (2006) Investigating emotion with music: an fMRI study. Human Brain Mapping, 27, 239-250.

Kostka, S., Payne, D., & Almen, B. (2012). Tonal Harmony (7 edition). New York: McGraw-Hill Education.

Kostka. (2003). Instructor's Manual to Accompany Tonal Harmony (5th Ed edition). Boston.

Krantz, S. E., & Rude, S. S. (1984). Depressive attributions: selection of different causes or assignment of dimensional meanings? Journal of Personality and Social Psychology, 47, 193-203.

Krumhansl, C. L. (2001). Cognitive Foundations of Musical Pitch. New York: Oxford University Press.

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. Psychophysiology, 30(3), 261–273. https://doi.org/10.1111/j.1469-8986.1993.tb03352.x

Lawson, R., Rees, G., & Friston, K. J. (2014) An aberrant precision account of autism. Frontiers in Human Neuroscience, 8, 302.

Lerdahl, F., & Krumhansl, C. L. (2007) Modeling tonal tension. Music Perception 24:329-366.

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: A meta-analytic review. Behavioral and Brain Sciences, 35(3), 121–143. https://doi.org/10.1017/S0140525X11000446

Ling, S., & Carrasco, M. (2006). When sustained attention impairs perception. Nature Neuroscience, 9(10), 1243–1245. https://doi.org/10.1038/nn1761

Loftus, G. R., & Masson, M. E. (1994) Using confidence intervals in within-subject designs. Psychonomic Bulletin & Review, 1, 476-490.

Losh, M., & Capps, L. (2006). Understanding of emotional experience in autism: insights from the personal accounts of high-functioning children with autism. Developmental Psychology, 42(5), 809–818. https://doi.org/10.1037/0012-1649.42.5.809

Lu, Z.-L., & Dosher, B. A. (1998). External noise distinguishes mechanisms of attention. Vision Res, 38, 1183–1198.

Maldjian, J. A., Laurienti, P. J., Kraft, R. A., & Burdette, J. H. (2003) An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fmri data sets. Neuroimage, 19, 1233-1239.

Malle, B. F. (2001) Folk explanations of intentional action. In B. F. Malle, L. J. Moses & D. A. Baldwin (Eds.), Intentions and intentionality: Foundations of social cognition (pp. 265-286). Cambridge, MA: MIT Press.

Mandler, J. M., & Johnson, N. S. (1976). Some of the thousand words a picture is worth. Journal of Experimental Psychology. Human Learning and Memory, 2(5), 529–540.

Marchetti, A., Baglio, F., Costantini, I., Dipasquale, O., Savazzi, F., Nemni, R., et al. (2015) Theory of Mind and the whole brain functional connectivity: behavioral and neural evidences with the Amsterdam Resting State Questionnaire. Frontiers in Psychology, 6, 1855.

Marco, E. J., Hinkley, L. B. N., Hill, S. S., & Nagarajan, S. S. (2011). Sensory Processing in Autism: A Review of Neurophysiologic Findings. Pediatric Research, 69(5 Pt 2), 48R–54R. https://doi.org/10.1203/PDR.0b013e3182130c54

Martin, A., & Weisberg, J. (2003) Neural foundations for understanding social and mechanical concepts. Cognitive Neuropsychology, 20, 575-87.

McAlonan, G. M., Cheung, C., Cheung, V., Wong, N., Suckling, J., & Chua, S. E. (2009) Differential effects on white-matter systems in high-functioning autism and Asperger's syndrome. Psychological Medicine, 39, 1885-1893.

McDermott, J. H., Lehr, A. J., & Oxenham, A. J. (2010). Individual Differences Reveal the Basis of Consonance. Current Biology, 20(11), 1035–1041. https://doi.org/10.1016/j.cub.2010.04.019

McDermott, J. H., Schultz, A. F., Undurraga, E. A., & Godoy, R. A. (2016). Indifference to dissonance in native Amazonians reveals cultural variation in music perception. Nature. https://doi.org/10.1038/nature18635

McDermott, J., & Hauser, M. (2004). Are consonant intervals music to their ears? Spontaneous acoustic preferences in a nonhuman primate. Cognition, 94(2), B11-21. https://doi.org/10.1016/j.cognition.2004.04.004

Menard, S. (1995) Applied logistic regression analysis: Sage University series on quantitative applications in the social sciences. Thousand Oaks: Sage.

Menon, V., & Uddin, L. Q. (2010) Saliency, switching, attention and control: a network model of insula function. Brain Structure and Function, 214, 655-667.

Meyer, L. B. (1956) Emotion and meaning in music. Chicago: Univ of Chicago Press.

Minati, L., Rosazza, C., D'Incerti, L., Pietrocini, E., Valentini, L., Scaioli, V., … Bruzzone, M. G. (2009). Functional MRI/event-related potential study of sensory consonance and dissonance in musicians and nonmusicians. Neuroreport, 20(1), 87–92. https://doi.org/10.1097/WNR.0b013e32831af235

Minsky, M. (1974). A framework for representing knowledge. Retrieved from https://dspace.mit.edu/handle/1721.1/6089

Mitchell, J. P. (2008) Activity in right temporo-parietal junction is not selective for theory-of-mind. Cerebral Cortex, 18, 262-271.

Mitchell, J. P., Banaji, M. R., & Macrae, C. N. (2005) The link between social cognition and self-referential thought in the medial prefrontal cortex. Journal of Cognitive Neuroscience, 17(8), 1306-15.
Mitchell, J. P., Cloutier, J., Banaji, M, R., & Macrae, C. N. (2006) Medial prefrontal dissociations during processing of trait diagnostic and nondiagnostic person information. Social Cognitive and Affective Neuroscience, 1, 49-55.
Mitterschiffthaler, M. T., Fu, C. H. Y., Dalton, J. A., Andrew, C. M., & Williams, S. C. R. (2007). A functional MRI study of happy and sad affective states induced by classical music. Human Brain Mapping, 28(11), 1150–1162. https://doi.org/10.1002/hbm.20337
Mobbs, D., Weiskopf, N., Lau, H. C., Featherstone, E., Dolan, R. J., & Frith, C. D. (2006). The Kuleshov Effect: the influence of contextual framing on emotional attributions. Social Cognitive and Affective Neuroscience, 1(2), 95–106. https://doi.org/10.1093/scan/nsl014
Montagne, B., Kessels, R. P. C., De Haan, E. H. F., & Perrett, D. I. (2007). The Emotion Recognition Task: A Paradigm to Measure the Perception of Facial Emotional Expressions at Different Intensities. Perceptual and Motor Skills, 104(2), 589–598. https://doi.org/10.2466/pms.104.2.589-598
Mottron, L., Dawson, M., Soulières, I., Hubert, B., & Burack, J. (2006). Enhanced Perceptual Functioning in Autism: An Update, and Eight Principles of Autistic Perception. Journal of Autism and Developmental Disorders, 36(1), 27–43. https://doi.org/10.1007/s10803-005-0040-7
Mueller, S., Keeser, D., Samson, A. C., Kirsch, V., Blautzik, J., Grothe, M., et al. (2013) Convergent findings of altered functional and structural brain connectivity in individuals with high functioning autism: a multimodal MRI Study. PLoS One, 8, e67329.
Mumford, D. (1992). On the computational architecture of the neocortex. II. The role of cortico-cortical loops. Biological Cybernetics, 66(3), 241–251.
Murray, S. O., Kersten, D., Olshausen, B. A., Schrater, P., & Woods, D. L. (2002) Shape perception reduces activity in human primary visual cortex. Proceedings of the National Academy of Sciences, U.S.A., 99, 15164-15169.
Myers. (1990) Classical and modern regression with applications. Boston: PWS-Kent.
Ozonoff, S., Strayer, D. L., McMahon, W. M., & Filloux, F. (1994). Executive Function Abilities in Autism and Tourette Syndrome: An Information Processing Approach. Journal of Child Psychology and Psychiatry, 35(6), 1015–1032. https://doi.org/10.1111/j.1469-7610.1994.tb01807.x
Parncutt, R. (1989) Harmony: a psychoacoustical approach. Berlin: Spring-Verlag.
Peelen, M. V., Atkinson, A. P., & Vuilleumier, P. (2010). Supramodal Representations of Perceived Emotions in the Human Brain. Journal of Neuroscience, 30(30), 10127–10134. https://doi.org/10.1523/JNEUROSCI.2161-10.2010
Pellicano, E, & Burr, D. (2012) When the world becomes "too real": a Bayesian explanation of autistic perception. Trends in Cognitive Sciences, 16, 504-510.
Peretz, I., Blood, A. J., Penhune, V., & Zatorre, R. (2001). Cortical deafness to dissonance. Brain: A Journal of Neurology, 124(Pt 5), 928–940.
Plomp, R., & Levelt, W. J. M. (1965) Tonal consonance and the critical bandwidth. The Journal of the Acoustical Society of America, 38, 548-560.
Rao, R. P., & Ballard, D. H. (1999) Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. Nature Neuroscience, 2, 79-87.
Rosch, E. (1975) Universals and cultural specifics in human categorization. In R. Breslin, S. Bochner, & W. Lonner (Eds), Cross-cultural perspectives on learning (pp. 117-206). New York: Halsted Press.
Russell, J. (Ed.). (1998). Autism as an Executive Disorder. Oxford, New York: Oxford University Press.
Russell, J. A. (1979). Affective space is bipolar. Journal of Personality and Social Psychology, 37(3), 345.
Russo, N. M., Skoe, E., Trommer, B., Nicol, T., Zecker, S., Bradlow, A., & Kraus, N. (2008). Deficient brainstem encoding of pitch in children with Autism Spectrum Disorders. Clinical Neurophysiology, 119(8), 1720–1731. https://doi.org/10.1016/j.clinph.2008.01.108
Rutherford, M. D., Baron-Cohen, S., & Wheelwright, S. (2002) Reading the mind in the voice: a study with normal adults and adults with Asperger syndrome and high functioning autism. Journal of Autism and Developmental Disorders, 32(3), 189-94.
Santiesteban, I., Banissy, M. J., Catmur, C., & Bird, G. (2012). Enhancing Social Ability by Stimulating Right Temporoparietal Junction. Current Biology, 22(23), 2274–2277. https://doi.org/10.1016/j.cub.2012.10.018
Saxe, R. (2006) Uniquely human social cognition. Current Opinion in Neurobiology, 16(2), 235-239.
Saxe, R. (2010) The right temporo-parietal junction: a specific brain region for thinking about thoughts. In A. Leslie, & T. German (Eds), Handbook of theory of mind (pp.1-35). Psychology Press.
Saxe, R., & Kanwisher, N. (2003) People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind". Neuroimage, 19, 1835-1842.
Saxe, R., & Wexler. A. (2005) Making sense of another mind: the role of the right temporo-parietal junction. Neuropsychologia, 43, 1391-1399.

Schank, R. C. (1975). Using knowledge to understand. In Proceedings of the 1975 workshop on Theoretical issues in natural language processing (pp. 117–121). Association for Computational Linguistics. Retrieved from http://dl.acm.org/citation.cfm?id=980223

Schellenberg, E. G, & Trehub, S. E. (1994) Frequency ratios and the discrimination of pure tone sequences. Perception & Psychophysics, 56, 472-478.

Schellenberg, E. G., & Trainor, L. J. (1996) Sensory consonance and the perceptual similarity of complex-tone harmonic intervals: tests of adult and infant listeners. Journal of the Acoustical Society of America, 100, 3321-3328.

Schellenberg, E. G., & Trehub, S. E. (1996a) Natural musical intervals: evidence from infant listeners. Psychological Science, 7, 272-277.

Schellenberg, E. G., & Trehub, S. E. (1996b) Children's discrimination of melodic intervals. Developmental Psychology, 32, 1039-1050.

Schellenberg, E. G., Peretz, I., & Vieillard, S. (2008). Liking for happy- and sad-sounding music: Effects of exposure. Cognition & Emotion, 22(2), 218–237. https://doi.org/10.1080/02699930701350753

Schilbach, L., Eickhoff, S. B., Rotarska- Jagiela, A., Fink, G. R., and Vogeley, K. (2008) Minds at rest? Social cog- nition as the default mode of cog- nizing and its putative relationship to the default system of the brain. Conscious. Cogn. 17, 457-467.

Scholz, J., Triantafyllou, C., Whitfield-Gabrieli, S., Brown, E. N., & Saxe, R. (2009) Distinct regions of right temporo-parietal junction are selective for theory of mind and exogenous attention. PLoS One, 4(3), e4869.

Schultz, J., Imamizu, H., Kawato, M., & Frith, C. D. (2004) Activation of the human superior temporal gyrus during observation of goal attribution by intentional objects. Journal of Cognitive Neuroscience, 16, 1695-705.

Schurz, M., & Perner, J. (2015) An evaluation of neurocognitive models of Theory of Mind. Frontiers in Psychology, 6, 1610.

Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., … Greicius, M. D. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 27(9), 2349–2356. https://doi.org/10.1523/JNEUROSCI.5587-06.2007

Seghier, M. L. (2013). The Angular Gyrus. The Neuroscientist, 19(1), 43–61. https://doi.org/10.1177/1073858412440596

Serafine, M. L. (1983) Cognition in music. Cognition, 14, 119-183.

Silani, G., Bird, G., Brindley, R., Singer, T., Frith, C., & Frith, U. (2008). Levels of emotional awareness and autism: an fMRI study. Social Neuroscience, 3(2), 97–112. https://doi.org/10.1080/17470910701577020

Silani, G., Lamm, C., Ruff, C. C., & Singer, T. (2013). Right Supramarginal Gyrus Is Crucial to Overcome Emotional Egocentricity Bias in Social Judgments. Journal of Neuroscience, 33(39), 15466–15476. https://doi.org/10.1523/JNEUROSCI.1488-13.2013

Soveri, A., Tallus, J., Laine, M., Nyberg, L., Bäckman, L., Hugdahl, K., … Hämäläinen, H. (2013). Modulation of Auditory Attention by Training. Experimental Psychology, 60(1), 44–52. https://doi.org/10.1027/1618-3169/a000172

Stephan, K. E., Penny, W. D., Moran, R. J., den Ouden, H. E., Daunizeau, J., & Friston, K. J. (2010). Ten simple rules for dynamic causal modeling. Neuroimage, 49, 3099-3109.

Sugimoto, T., Kobayashi, H., Nobuyoshi, N., Kiriyama, Y., Takeshita, H., Nakamura, T., & Hashiya, K. (2010). Preference for consonant music over dissonant music by an infant chimpanzee. Primates; Journal of Primatology, 51(1), 7–12. https://doi.org/10.1007/s10329-009-0160-3

Temperley, D. (2007) Music and probability. Cambridge: MIT Press.

Terhardt, E. (1978) Psychoacoustic evaluation of musical sounds. Perception & Psychophysics, 23, 483-492.

Terhardt, E. (1984) The concept of musical consonance: a link between music and psychoacoustics. Music Perception, 1, 276-29.

Thakkar, K. N., Polli, F. E., Joseph, R. M., Tuch, D. S., Hadjikhani, N., Barton, J. J., et al. (2008) Response monitoring, repetitive behaviour and anterior cingulate abnormalities in autism spectrum disorders (ASD). Brain, 131, 2464-2478.

Todorov, A., Gobbini, M. I., Evans, K. K., & Haxby, J. V. (2007) Spontaneous retrieval of affective person knowledge in face perception. Neuropsychologia, 45, 163–73.

Trainor, L. J., & Heinmiller, B. M. (1998) The development of evaluative responses to music: infants prefer to listen to consonance over dissonance. Infant Behavior and Development, 21, 77-88.

Trehub, S. E., & Unyk, A. M. (1991) Music prototypes in developmental perspective. Psychomusicology, 10, 31-45.

Uddin, L. Q. (2015) Salience processing and insular cortical function and dysfunction. Nature Reviews Neuroscience, 16, 55-61.

Uljarevic, M., & Hamilton, A. (2012) Recognition of emotions in autism: a formal meta-analysis. Journal of Autism and Developmental Disorders, 43(7), 1517-1526.

Van Overwalle, F. (2009) Social cognition and the brain: a meta-analysis. Human Brain Mapping, 30(3), 829-858.

Van Overwalle, F., & Baetens, K. (2009) Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis. Neuroimage, 48, 564-84.

Van Overwalle, F., Van den Eede, S., Baetens, K., & Vandekerckhove, M. (2009). Trait inferences in goal-directed behavior: ERP timing and localization under spontaneous and intentional processing. Social Cognitive and Affective Neuroscience, 4(2), 177–190. https://doi.org/10.1093/scan/nsp003

Van Overwalle, F., Van den Eede, S., Baetens, K., & Vandekerckhove, M. (2009) Trait inferences in goal-directed behavior: ERP timing and localization under spontaneous and intentional processing. Social Cognitive and Affective Neuroscience, 4, 177-190.

Vos, J., & Vianen, B. G. (1984) Thresholds for discrimination between pure and tempered intervals: the relevance of nearly coinciding harmonics. The Journal of the Acoustical Society of America, 77, 176-187.

Wedin, L. (1972) A multidimensional study of perceptual-emotional qualities in music. Scandinavian Journal of Psychology, 13, 241-257.

Williams, D., & Happé, F. (2010). Recognising “social” and “non-social” emotions in self and others: A study of autism. Autism, 14(4), 285–304. https://doi.org/10.1177/1362361309344849

Woodbury-Smith, M. R., Robinson, J., Wheelwright, S., & Baron-Cohen, S. (2005) Screening adults for Asperger syndrome using the AQ: a preliminary study of its diagnostic validity in clinical practice. Journal of Autism and Developmental Disorders, 35, 331-335.

Young, L., & Saxe, R. (2008) The neural basis of belief encoding and integration in moral judgment. Neuroimage, 40, 1912-1920.

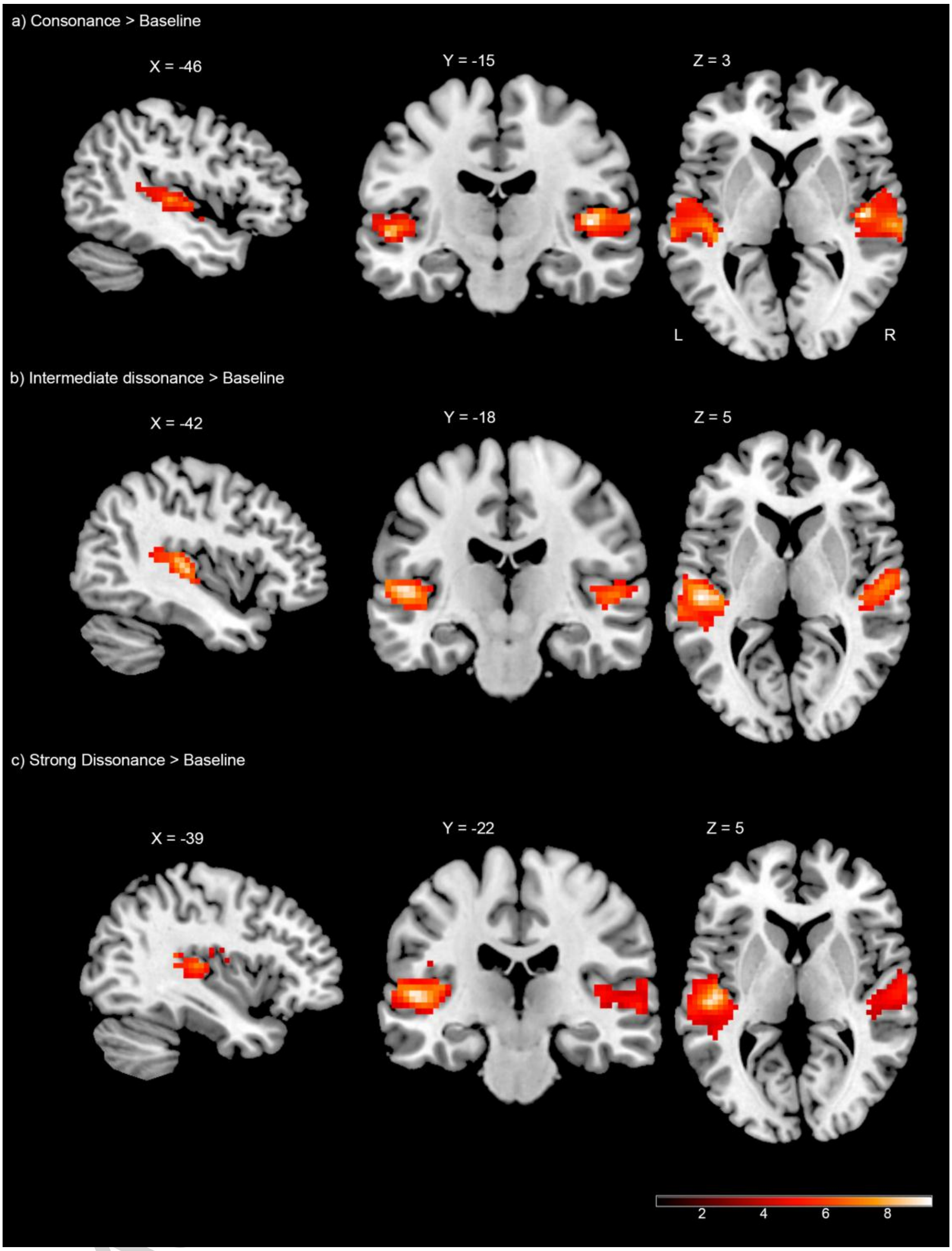
Young, L., & Saxe, R. (2009a) Innocent intentions: A correlation between forgiveness for accidental harm and neural activity. Neuropsychologia, 47, 2065-2072.

Young, L., & Saxe, R. (2009b) An FMRI investigation of spontaneous mental state inference for moral judgment. Journal of Cognitive Neuroscience, 21, 1396-1405.

Young, L., Camprodon, J. A., Hauser, M., Pascual-Leone, A., & Saxe, R. (2010) Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. Proceedings of the National Academy of Sciences, U.S.A., 107, 6753-6758.

Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007) The neural basis of the interaction between theory of mind and moral judgment. Proceedings of the National Academy of Sciences, U.S.A., 104, 8235-8240.

Zentner, M. R., & Kagan, J. (1996) Perception of music by infants. Nature, 383, 29.



Supplementary Figure S5. FMRI results (FWE-corrected P < 0.05 for cluster-level inference, whole-brain analysis). Statistical parametric maps (SPM) of the direct contrast between sound and baseline conditions superimposed onto a standard brain in stereotactic MNI space. Stronger BOLD signals were found in bilateral superior temporal regions, including Heschl's gyri (primary auditory cortex) (Table 5A).

Supplementary Table S8. Internal AQ sub-scores (social skills, attention switching, attention to detail, communication and imagination), total AQ scores and gender for the thirty-nine individuals that participated in the internet-based experiment.

Participant number	Social skills	Attention switching	Attention detail	Communication	Imagination	Total AQ	Gender
1	3	6	7	1	3	20	M
2	1	7	4	1	2	15	M
3	6	6	5	3	3	23	M
4	1	5	6	2	4	18	M
5	2	4	2	0	2	10	M
6	7	7	3	3	4	24	M
7	1	3	4	1	4	13	M
8	1	5	4	2	2	14	M
9	4	4	7	2	3	20	M
10	3	7	9	4	4	27	M
11	2	2	4	0	1	9	M
12	8	6	2	6	8	30	M
13	1	6	10	2	3	22	M
14	5	4	2	3	3	17	M
15	2	3	1	0	2	8	M
16	3	4	6	1	3	17	M
17	1	7	8	1	2	19	M
18	2	5	7	4	3	21	M
19	2	3	5	2	1	13	M
20	1	3	6	0	2	12	M
21	3	5	8	1	1	18	F
22	4	6	6	4	6	26	F
23	10	9	7	4	3	33	F
24	4	2	7	3	3	19	F
25	3	5	4	3	3	18	F
26	5	6	4	3	3	21	F
27	1	4	5	1	1	12	F
28	6	8	5	1	5	25	F
29	1	6	7	2	4	20	F
30	6	3	2	5	6	22	F
31	1	3	3	1	3	11	F
32	2	5	3	0	5	15	F
33	3	2	7	0	5	17	F
34	2	7	4	1	2	16	F
35	2	5	6	2	6	21	F
36	7	6	4	3	6	26	F
37	2	4	6	3	5	20	F
38	4	6	5	1	4	20	F
39	2	6	4	2	3	17	F

Highlights

- Subjects made valence inferences of stimuli with different levels of dissonance.
- Consonance/dissonance degree modulated responses on the right angular gyrus.
- Dissonance demanded heightened ‘theory of mind’ and attention reorienting resources.
- Social skills traits associated with deficits in valence appraisal of consonances.